PREDICTION OF SPATIAL-TEMPORAL DISTRIBUTION OF ALGAL METABOLITES IN EAGLE CREEK RESERVOIR, INDIANAPOLIS, IN

Slawa Romana Bruder

Submitted to the faculty of the University Graduate School
in partial fulfillment of the requirements
for the degree
Master of Science
in the Department of Earth Sciences,
Indiana University

May 2012

Accepted by the Faculty of Indiana University, in partial fulfillment of the requirements for the degree of Master of Science.

Meghna Babbar-Sebens Ph.D., Chair

Pierre-Andre Jacinthe Ph.D.

Master's Thesis Committee

Lenore P. Tedesco Ph.D.

ACKNOWLEDGEMENTS

I would like to express my deepest gratitude toward the research advisor Dr. Meghna Babbar-Sebens for accepting me as part of her research team. Her belief in one's intellectual potential and continuous enthusiasm and support were very inspiring throughout the process of acquiring new experiences and allowed me to reach higher levels of professionalism.

I would also like to thank research committee members Dr. Lenore Tedesco and Dr. Pierre-Andre Jacinthe for their support and scientific contributions. I am thankful to Dr. Tedesco for her dedication and true knowledge of environmental sciences. I am thankful to Dr. Jacinthe for his positive demeanor and inquisitive nature.

A very special thank you goes to my family, whose love and support are independent of passing time. I am thankful to my parents, Danuta and Zbigniew Zań, whose ability to see wonder in everything from big to small and whose dedication to helping those in need, will always guide me in my pursuit of a fulfilled and moral life. Another thank you goes to my husband Neal Bruder for walking with me the path of life, for unconditional love, support and knowledge. Also, thank you to my children, Maximillian and Charles for giving me the privilege to be a mother. I send my thank you to other members of my family: Marty, Ron Bruder, Carolyn Rodgers, and Elizabeth Kratz for encouragement, help, and spontaneous love. Along with my family, I would like to thank Shuangshuang Xie for giving me a pleasure of knowing someone with such a beautiful, admirable nature, who has not only been my colleague, friend but many times a teacher.

I would like to thank Nicholas Clercin and Michael Stouder for their contribution to this research and eagerly shared knowledge.

Finally, I would like to thank the Central Indiana Water Resources Partnership (CIWRP) and Veolia Water Indianapolis, LLC. for project funding and the Center for Earth and Environmental Science at IUPUI for logistical support and background data.

ABSTRACT

Slawa Romana Bruder

PREDICTION OF SPATIAL-TEMPORAL DISTRIBUTION OF ALGAL METABOLITES IN EAGLE CREEK RESERVOIR, INDIANAPOLIS, IN

In this research, Environmental Fluid Dynamic Code (EFDC) and Adaptive- Network-based Fuzzy Inference System Models (ANFIS) were developed and implemented to determine the spatial-temporal distribution of cyanobacterial metabolites: 2-MIB and geosmin, in Eagle Creek Reservoir, IN. The research is based on the current need for understanding algae dynamics and developing prediction methods for algal taste and odor release events.

In this research the methodology for prediction of 2-MIB and geosmin production was explored. The approach incorporated a combination of numerical and heuristic modeling to show its capabilities in prediction of cyanobacteria metabolites. The reservoir's variable data measured at monitoring stations and consisting of chemical/physical and biological parameters with the addition of calculated mixing conditions within the reservoir were used to train and validate the models. The Adaptive – Network based Fuzzy Inference System performed satisfactorily in predicting the metabolites, in spite of multiple model constraints. The predictions followed the generally observed trends of algal metabolites during the three seasons over three years (2008-2010). The randomly selected data pairs for geosmin for validation achieved coefficient of determination of 0.78, while 2-MIB validation was not accepted due to large differences between two observations and their model prediction. Although, these ANFIS results were accepted, the further application of the ANFIS model coupled with the numerical models to predict spatio-temporal distribution of metabolites showed serious limitations, due to numerical model calibration errors. The EFDC-ANFIS model over-predicted *Pseudanabaena spp*. biovolumes for selected stations. The predicted value was 18,386,540 mm^3/m^3, while observed values were 942,478 mm^3/m^3. The model simulating *Planktothrix agardhii* gave negative biovolumes, which were assumed to represent zero values observed at the station. The taste and odor metabolite, geosmin, was under-predicted as the predicted

concentration was 3.43 ng/L in comparison to observed value of 11.35 ng/l. The 2-MIB model did not validate during EFDC to ANFIS model evaluation.

The proposed approach and developed methodology could be used for future applications if the limitations are appropriately addressed.

Meghna Babbar-Sebens Ph.D., Chair

CONTENTS

List of Tables

List of Figures

1. Introduction

Occurrences of excessive algal blooms in water bodies worldwide are indicators of increasing water quality problems resulting from nutrient over-enrichment, modified hydrology, and poor watershed management. Presence of algal blooms is only one of the signs of eutrophication and is interconnected with the occurrence of other events such as dissolved oxygen depletion, low water transparency, development of nuisance or exotic animal population, fish kills, impaired potable water supplies, tainted fish and noxious odors. These environmental outcomes lead to further economic losses, which manifest themselves in decreased property values, high costs of raw drinking water, illnesses and depressed recreational industries. For example, the potential annual value losses in waterfront real estate, recreational water usage, spending on recovery of threatened and endangered species and drinking water due to eutrophication have been estimated to be 2.2 billion annually for U.S. freshwaters (Dodds et al., 2009).

The growing need for better management of algal blooms and their metabolites has led to the development of multiple heuristic and mechanistic algal bloom prediction models that incorporate the relationships between physical variables, chemical composition and ecological dynamics of algal blooms in the aquatic environment (U.S. EPA, New England Region, 2002; Marsili-Libelli, 2004; Tillman et al., 2004; Blauw et al., 2006; Chen et al., 2006; Velo-Suárez, 2007; Tetra Tech, Inc. 2009). Numerical models have shown great performance in mechanistic modeling of physical processes, but are still premature when it comes to simulating ecological and biological responses of living organisms whose underlying processes are often unknown. On the other hand, data-based heuristic models can benefit from modeler's knowledge and monitored data to incorporate different degrees of unknown relationships between variables. For example, different forms of heuristic models have been used to simulate algal mass occurrences, harmful algal blooms (Blauw et al., 2006; Malve et al., 2007; Velo-Suárez, Gutiérrez-Estrada, 2007). Malve et al. (2007) used bayesian modeling and Markov chain Monte Carlo simulation in an attempt to capture the response of phytoplankton to variation in nutrients (phosphorus, total nitrogen), water temperature, irradiance and total zooplankton biomass. Blauw et al. (2006) used fuzzy logic to model blooms of

Alexandrium minutum, Nodularia spumigena, Dinophysis spp., Karenia mikimotoi and Phaeocystis globosa. Velo-Suárez and Gutiérrez-Estrada (2007) applied artificial neural networks (ANN) in prediction of *Dinophysis acuminata* blooms in coastal waters of southern Spain.

Proven capabilities of these two different modeling approaches (i.e. mechanistic and data-based heuristic modeling approaches) create the opportunity for integrating multiple models that define the dependency between the physical, chemical parameters, and biological parameters (e.g., algal species, and algal metabolites) in the water column. Existing research shows that blooms of algal functional groups (Nielsen, 2005; Galloway, Green, 2006; McGovern, 2006) and species (Recknagel, 1997; Blauw et al., 2006) have been successfully modeled with inclusion of site specific data. However, fate and transport of algal metabolites, such as geosmin and 2 –Methylisoborneol (2-MIB) produced by certain cyanobacteria species, have not been modeled due to limited understanding of underlying processes. The production of metabolites is species and strain specific, which means that only selected strains of cyanobacteria in a given species can produce these compounds while others cannot (Taylor et al., 2006). Combined with environmental parameters, they represent a complex system comprised of multiple interactions and non-linear behaviors. The example of these interactions includes the species competition, which can lower particular species' abundance. The species competition, in turn, is determined by many factors including nutrient availability, presence of grazing zooplankton and physical parameters (temperature, light availability, mixing). The cyanobacteria may respond to these inter-dependent environmental stressors by producing algal metabolites (geosmin, 2-MIB).

In addition, as indicated by previous studies (Ligor, Buszewski, 2005; Uwins et al., 2007; Winston, 2010), there can be multiple possible sources of geosmin and MIB in the water column (e.g., standing timber, vegetation, actinomycetes, etc.) and the detection of these compounds can be hindered by the presence of dissolved lipids and humic acids. This introduces substantial uncertainty in determining the metabolite-algae relationships. Also, geosmin and 2-MIB can also be accidental by-products of an organism trying to sequester

phosphate to maintain the metabolic processes (Uwins et al., 2007). Considering all the factors, data-based heuristic models that simulate observed species behavior in an aquatic system, with no rigid assumptions on the multiple confounding drivers and variables, can pose as better alternatives for combining both quantitative and qualitative knowledge of ecological and physical processes with some degree of certainty. Existing applications that have used data-based approaches, such as Fuzzy Logic Models and artificial network approaches (Jang, Sun, 1995; Blauw et al., 2006; Chen et al., 2006; Velo-Suárez, 2007) as possible heuristic models for simulating algae dynamics at species levels have found that expert knowledge on local conditions and local interactions can play a significant role in identifying dynamics at this level. There has, however, been limited previous research in developing predictive models for taste and odor (T&O) compounds released by micro-algal communities. Most of the previous quantitative models have been regression based (Dzialowski et al., 2007; Wyrobek, 2010) and have not incorporated the uncertainty in multiple confounding variables.

The overarching goal of this study was to investigate whether an integrated modeling approach can be used to incorporate available physical and chemical data for accurately predicting the spatial-temporal distribution of algal blooms, and their taste and odor by-products in a freshwater inland drinking water supply reservoir – Eagle Creek Reservoir, IN. An integrated modeling approach that combined mechanistic models (Soil and Water Assessment Tool - SWAT, Environmental Fluid Dynamic Code - EFDC, and Hydrodynamic-Eutrophication Model - HEM-3D) with an empirical fuzzy logic model (Adaptive-Network-based Fuzzy Inference System (ANFIS)) has been developed in this research. The outcome of this research was to provide multivariate relationships between nutrient composition, presence of particular algal species, and taste and odor production within the reservoir at a given spatial-temporal scale, in order to develop a quantitative prediction system for the taste and odor algal metabolites in a particular reservoir. Specific objectives of this research were:

A. Development of mechanistic models for quantifying spatial-temporal distribution of micro-algae functional groups and nutrients in Eagle Creek Reservoir in Indianapolis, IN

using Environmental Fluid Dynamic Code (EFDC) and Hydrodynamic-Eutrophication Model (HEM-3D) for 2008 data.

B. Development of an empirical Adaptive-Network-based Fuzzy Inference System model that relates taste and odor (i.e., 2-MIB and geosmin) release events to physical, chemical, and biological conditions in the reservoir over the period 2008-2010.

C. Prediction of spatial–temporal distribution of concentrations of 2-MIB and geosmin occurrence by coupling models developed under objectives A and B.

2. Background

Complexity of water-algae interactions requires careful analysis of all environmental (physical, chemical and biological) components as they change with each site and system. Additionally, selection of correct parameters/variables is critical in successfully correlating by-products with algal blooms. The complexity of parameter interactions often causes the modeling matrix to be comprised of an excessively large number of variables, hence, increasing the risk of developing an over-fitted model. To overcome this hurdle many researchers have tried to identify the most significant factors/parameters that influence bloom occurrence, and thereby remove redundant parameters from their models or/and monitoring studies (Wong et al., 2009; Marsili-Libelli, 2004). Parameters such as: dissolved oxygen, oxidation-reduction potential, pH, temperature, light extinction, nutrients, tidal flushing, wind and tidal mixing are among the most commonly mentioned algal bloom factors. Dzialowski et al. (2007) combined water quality modeling techniques with landscape models using remotely sensed data and GIS to predict T&O events. Their study indicated that trophic state wasn't a good taste and odor predictor and that individual (i.e. specific to a site) rather than universal modeling approach was more appropriate for these kinds of investigations. Phosphorus, total cyanobacteria biovolume and sechii disk depth were found to be strong geosmin predictors. It has also been observed in previous studies (Ligor, Buszewski, 2005) that concentrations of 2-MIB and geosmin can be strongly correlated to annual seasons, and changes in geosmin concentration can be linked to water temperature, pH and conductivity. Yet, other researchers (Journey et al., 2008) correlated algal blooms with geosmin concentration levels and showed that (a) established nutrient criteria didn't correlate with phytoplankton community structure and (b) geosmin levels could be high during de-stratified conditions. Journey et al. (2008) also reported in their study that they did not observe any pattern between cell density, geosmin producing cyanobacteria, and geosmin occurrence. It was concluded that elevated concentrations of geosmin were complexly interrelated with nutrient dynamics (concentrations of inorganic nitrogen), type and density of cyanobacterial species, water temperature and degree of stratification. In some cases, investigations have clearly shown dependence of odor metabolite release on the density of cyanobacterial cells (Hobson et al. 2010). While Hobson et al. (2010) related

non detection of algal metabolites to loss factors such as biodegradation and volatilization, others (Zhang et al., 2009) have indicated that geosmin and 2-MIB are relatively stable to chemical and biological degradation. A study by Zhang et al. (2009) focused on temperature and light influence on metabolites and pointed out that low temperature and light could simulate geosmin production and favor the accumulation of geosmin in cells, while a higher amount of the intracellular geosmin may be released at optimum light intensity or higher temperatures. The other approaches in taste and odor monitoring for early warning programs have included study of log and lag growth phases of the metabolite producing algal blooms to capture early T&O outbreaks (Taylor et al., 2006). These and other studies underscore the complexity and uncertainty in developing quantitative relationships between driving factors and T&O compounds, and the need for developing site-specific models for simulating production of T&O compounds in a particular aquatic system.

3. Case Study

Eagle Creek Reservoir (ECR) is located northwest of Indianapolis, IN and lies within the Eagle Creek watershed, which covers the area of 426 km ^{2} (Figure 1). The Reservoir was constructed in 1967 for flood control and later modified for use as a drinking water supply reservoir for Indianapolis. The total size of the reservoir is 5.1 km^2 with mean depth measured at 5.7 m. Eagle Creek Reservoir can be characterized into three areas: northern basin (2.07 km^2), southern basin (2.97 km^2), and quarry (0.57 km^2). The 56th Street Bridge separates the north and south basins creating a limited water exchange passage with a 50 m opening (Figure 2). The quarry has negligible surface water interactions with the northern and southern basins.

The reservoir is divided into three zones based on physical characteristics: riverine (consisting of narrow, well mixed zone with river-like flow conditions – in the northern basin), transitional (transition from river-like to lacustrine zone - 56th Street Bridge area) and lacustrine (consisting of buoyancy forces that dominate the flow patterns in the southern basin). Normal pool elevation for ECR is 240.79m above sea level. There are two structures within the reservoir: dam and water intake. Approximately 10 million gallons per day (MGD) of water is pumped by the drinking water utility company (Lobugeois, 2009) via six gates located at three levels: two gates at 234.01 m, two gates at 236.14 m and two gates at 238.27 m (i.e. the water intake).

The major streams/tributaries entering the Eagle Creek reservoir are Eagle Creek, Bush Creek, Fishback Creek, and School Branch, though several smaller coves also exist on the boundary of the reservoir, many of which have small, ephemeral streams draining to the reservoir. These streams flowing through the watershed, which according to 2007 land use assessment was 45 % corn and soybean row crop agriculture, 18 % herbaceous land cover reflecting pasture and suburban land use, 13 % forested and 19 % urban development (Tedesco, personal communication) negatively affect the water quality of the reservoir due to their impairments including sedimentation, nutrient enrichment, toxic substances, and low oxygen concentrations (Bright, Cutler, 2002).

Eagle Creek Reservoir is a major source of drinking water to more than 800,000 (2010 United States Census) people living in the metropolitan city of Indianapolis, and also serves as an important venue for recreational activities in the area. In recent years algal bloom events have increased in the reservoir due to increased nutrients causing multiple water management concerns.

4. Methodology

4.1. General

The numerical models Environmental Fluid Dynamic Code (EFDC) and HEM-3D were used to simulate hydrodynamics and water quality of Eagle Creek Reservoir in Indianapolis, IN. Artificial Neuro-Fuzzy Inference System (ANFIS) was applied to the biological, physical and chemical data in an attempt to correlate reservoir's nutrient and algal composition to cyanobacteria species and their taste and odor components: 2-methyl isoborneal (2-MIB), geosmin. The general conceptual model is presented in Figure 3. The rectangle nodes confined in the larger rectangle blocks represent the models and the oval nodes the model's output. The overall modeling output is represented by dashed - outlined oval nodes. The arrows represent the direction of the modeling procedures.

The modeling framework followed a set of procedures. The collected field measured data, including climatic data from National Climate Data Center (NCDC) and Indiana State Climate Office for 2008, were combined with a watershed models (i.e. SWAT) output values for flow as inputs into EFDC hydrodynamic model. The calibrated hydrodynamic model was next imported to EFDC water quality module - HEM 3D. The HEM 3-D combined the calibrated hydrodynamic model variables, Soil and Water Assesment Tool (SWAT) output for nutrients in tributary inflows, chemical and biological data in the reservoir collected by Center for Earth and Environmental Sciences, and climatic data from NCDC. Additionally, literature based constants for nutrient groups (carbon, phosphorus, nitrogen), oxygen, algal stoichiometry and temperature were defined and used during model set up. The carbon, phosphorus and nitrogen nutrient groups were partitioned into refractory and labile fractions. The output from HEM-3D produced spatial-temporal distributions of three functional algal groups (cyanobacteria, diatoms, green algae) and nutrient concentrations. These results combined with additional field data (atmospheric precipitation, and water column stability) were used as an input to the ANFIS 1 and ANFIS 2 models. The ANFIS 1 model was designed to predict the cyanobacteria species producing T&O in the reservoir: *Pseudanabaena spp*. and *Planktrothrix agardhii*. The predicted cyanobacteria abundance was used as inputs, with other physical and chemical inputs, to the ANFIS 2 model. The ANFIS 2 model was

divided into two sub-models: one sub-model for predicting geosmin and one sub-model for predicting 2-MIB. The overall output combined the outputs of in EFDC and ANFIS models to predict the spatial and temporal distribution of algal metabolites in the reservoir.

4.2. Data Analysis and Preprocessing

4.2.1. EFDC Model Data

Data used in the proposed project comes from several sources including Center for Earth and Environmental Sciences (CEES), National Climatic Data Center, Indiana State Climate Office, and USGS monitoring stations.

Atmospheric Data

Hourly atmospheric data was obtained from the National Climatic Data Center (http://www.ncdc.noaa.gov/). Observations came from Eagle Creek Airpark (station ID 53842) station and included: wet/dry air bulb temperature, precipitation, relative humidity, air pressure, wind speed and direction, and cloud cover (data is reported in SI units).

Hourly solar radiation for 2008 was obtained from the Indiana State Climate Office (http://climate.agry.purdue.edu/climate/). The closest station - Throckmorton (TPAC) - reporting solar radiation is located in Lafayette, Tippecanoe County, IN which is 63 miles Northwest of Indianapolis. TPAC is the only data source for solar radiation which is available, but it belongs to a different ecoregion. Also, radiation at TPAC is measured by a LICOR 200 pyranometer sensor, which does not cover the entire short wave spectrum, only partially satisfied the model requirement (Figure 4). The solar radiation imported to the EFDC as boundary condition highly influences the model output. The EFDC is capable of internally calculating the solar radiation, therefore, two simulations, one of which uses TPAC data and one of which internally calculates solar radiation were run to check for discrepancies and sensitivity of the model to the solar radiation data. Wet and dry atmospheric deposition data for ammonium and nitrate were imported from National Atmospheric Deposition Program (http://nadp.sws.uiuc.edu/). The reporting station for

atmospheric deposition data - Agronomy Center for Research and Extension - is located in West Lafayette, IN at Purdue University.

An average estimated evaporation rate of 5.50 mm/day for the period of June to October, provided by CEES, was used. The averaged value is based on daily evaporation measurements taken in Carmel (27 km east from Eagle Creek Reservoir) by Veolia Water Company from June to October, 2008 (Lobugeois, 2009).

Bathymetry Data

Bathymetry data for Eagle Creek Reservoir was based on the measurements made by the Indiana Department of Natural Resources in 1994 (IDNR) using a Raytheon Echo Sounder –Model 719 C. Due to insufficient data under the 56th Street Bridge, more measurements were conducted for this area on June 22, 2009 by USGS. Accuracy of sonar used in 1994 and 2009 are estimated to be +/- 10 cm.

Watershed/Reservoir Data

Inflow data for the reservoir was estimated from a distributed hydrologic model based on the Soil and Water Assessment Tool (SWAT), due to lack of monitoring data for inflows. Based on the SWAT sub-basin division (Figure 5), 12 tributary inflows were specified for Eagle Creek Reservoir. There are two outflows from Eagle Creek Reservoir: Eagle Creek Reservoir dam and the water intake. Flow data at USGS gauge station # 03353460 at Eagle Creek, Clermont, IN (http://waterdata.usgs.gov/) was used as dam release information. Also, daily water intake data was obtained from the T. W. Moses drinking water facility (located within the reservoir).

Daily pool elevation data for 2008 came from USGS gauge station # 03353450 in the Eagle Creek Reservoir located in the eastern part of the dam (Figure 6).

Water quality data including physical, chemical, and biological properties was obtained from CEES and was based on measurements performed from May to October of 2008. Water quality data was collected in four main regions of the reservoir: near the Eagle

Creek tributary inflow (A), near the marina (B), near the water intake (C) and near the dam (D) (Figure 7). There were total of 54 monitoring locations for physical parameters within which 24 stations recorded chemical/biological data for 2008 (A-N). Each station with the unique geographic location reported using UTM coordinates was sampled only once. Many of the coordinates recorded via a global positioning system (GPS) device for the sampling locations in the dam, marina, and water intake regions were assessed to be inaccurate due to possible human errors. Hence, after communication with the data collection team, it was decided to collapse these sampling locations into representative locations for the regions, based on the information of the landmarks that the data collection team used to ensure collection at the representative location. The three main representative sampling locations (in UTM coordinates) in the dam, marina and intake regions provided by the data collection team are as follow: 559629-easting, 4408496-northing (dam), 559246-easting, 4411179-northing (intake), and 559252-easting, 4412838-northing (marina). These three locations (region A combined with B) were used in water quality modeling calibration instead of 54 locations originally indicated by GPS records. The monitored parameters contained: temperature, total dissolved solids (TDS), salinity, dissolved oxygen (DO), pH, Secchi depth, photic depth, silica (Si), total and dissolved organic carbon (DOC), nitrate + nitrite (NO_3+NO_2), total nitrogen (TN), ammonium (NH_4), total Kjeldahl nitrogen (TKN), total phosphorus (TP), total suspended sediment (TSS), chlorophyll-a, cyanobacteria, diatoms, green algae, geosmin and 2-MIB. Since water temperature for Eagle Creek Reservoir was not collected during winter and early spring in 2008, data from the nearest USGS monitoring stations was used for estimating the model's initial and boundary conditions for temperature (Figure 8). There were only two stations, which provided complete data for water temperature for 2008: USGS gauge station # 03359000 at Mill Creek (located in a different ecoregion) and USGS gauge station # 03354000 at White River near Centerton, IN (approximately 25 miles south from ECR). Water temperature for tributaries from January 1 to January 14, 2008 was estimated based on linear relationship between USGS Centerton station and Mill Creek station as Centerton station records didn't include fourteen days in January (Figure 9). Remaining values for tributary boundary conditions came from USGS Centerton station, since it is located relatively closer than the other USGS gauge. Nearest

gauge to reservoir, USGS Stout Gen. Station # 03353611 was not used because it was reported to be thermally polluted. The decision to use Centerton water temperature measurements for 2008 was validated using comparisons of available measurements for 2011 for Zionsville USGS gage 3353200 (located in the Eagle Creek watershed) and USGS gage 3354000 near Centerton (Figure 10). The 2011 data was used for this validation due to lack of 2008 values for Zionsville station. The average difference between these two stations for period July 14 to August 23 was 2.2 degrees Celsius and R^2 (trendline based) was 0.83 (Figure 11), while coefficient of determination of R^2 = -0.33 indicated that the average value of Zionsville data is better predictor than Centerton data. Data was accepted with the inclusion of existing difference.

4.2.2. ANFIS Model Data

Data for ANFIS models was provided by CEES (IUPUI). It consisted of photic depth averaged physical, chemical water parameters and algal species bio-volume records collected during 2008, 2009 and 2010 at monitoring stations located within ECR (Tables 1-3). The decision to incorporate 3 years of algal data (2008, 2009, and 2010), instead of using just the 2008 water quality dataset that was used for the EFDC-HEM3D models, was based on preliminary model simulations that indicated the 2008 dataset to be insufficient. In additional to field data, fifteen (15) day precipitation moving average, diffusion coefficient, calculations of relative thermal resistance to mixing (RTRM) and estimation of Richardson numbers (R_i) were also included in the data set. RTRM and R_i were used to estimate mixing conditions in the reservoir at the time the other water quality data were collected. There were a total of 112 actual locations (or, stations) from which water quality data was used in modeling (20 stations in 2008, 45 stations in 2009 and 44 stations in 2010). Algal data consisted of planktonic taxa and did not incorporate benthic populations. Since, no data for benthic algal species and taste and odor producing actinomycete soil bacteria was available, interpretation of species contribution to T&O was limited. The ANFIS model targeted selected cyanobacterial species present in ECR and recognized *Pseudanabeana spp.* and *Planktrothrix agardhii* as T&O producers. The selection of the two T&O producing species was based on communication with CEES representatives, Dr. Lenore Tedesco and Nicolas Clercin and their algal dynamics

knowledge of ECR in 2008, which indicated that the two described species were possible source of T&O at the time. This assumption didn't account for the other possible sources as soil bacteria and species, which at this time are not indicated as producers in existing literature.

4.2.3. Data Pre-processing

EFDC WQ Model

The methodology for estimating the labile, refractory and dissolved fractions of nitrogen and phosphorus for boundary and initial conditions was based on Christina River Basin modeling report (USEPA, 2000), since preliminary work on partitioning the nutrients has not been developed for Eagle Creek watershed. The Christina River Basin is located in Delaware. The climate of the Delaware state is modified humid continental with monthly average temperature of 12.7 ^{0}C (degrees Celsius). Indiana's climate is humid continental with monthly average temperatures between 9-12 ^{0}C in the north and 14 ^{0}C in the south. The land use described in USEPA modeling document (USEPA, 2000) for Christina River Sub-basin include 25 % residential use, 22 % urban, 22 forested, 12 % agricultural use, 12 % open and 7 % other. The dissolved organic phosphorus and nitrogen were estimated to be 0.50 fraction of total organic constituent, with each of labile and refractory particulate fractions being 0.25 of the total (USEPA, 2005). There two case studies share the climate similarities, however the land use statistics indicate, that the ECW contains more agriculture and not as much forested buffer in comparison with Christina River Basin, while The Christina River Basin has more urban areas.

Due to a lack of biochemical oxygen demand (BOD) and carbonaceous BOD (CBOD) data, the carbonaceous material in each of the tributary inflows were estimated based on sparse watershed averaged values of total organic carbon (TOC) and dissolved organic carbon (DOC).The TOC and DOC values collected within the EC watershed during each month were averaged and the refractory and labile particulate organic carbon was estimated to be fifty percent of the difference between total organic carbon and its dissolved fraction. These estimates were uniformly applied to all the tributaries.

The available (dissolved) and particulate biogenic silica (produced through diatom mortality) was assigned as default value, of 1 mg/l and 0.1 mg/l respectively and uniformly applied to tributary inflows. The available sparse measurement of silica concentrations in Eagle Creek watershed averaged for the 2008 was 6 mg/L. However, the EFDC boundary conditions inputs required the daily silica values and these input values were decreased as part of the modeler's decision making.

As indicated by Cerco (2000), when there are no observations available, then the feasible range is determined by parameter values used in similar models or by the judgment of the modeler.

Algal carbon content per unit species expressed as picogram per cell (pg/cell) was provided by CEES. These values were multiplied by algal cell counts in 1 milliliter of solution (cell/ml) to obtain the concentration of carbon in the solution of cells in pg/ml units. Then, a conversion factor was used to obtain the concentration in final units of mg/L.

ANFIS Model

In-situ algal data collected in the reservoir consisted of a composite sample obtained by sampling over the photic depth. Therefore, samples of physical variables (e.g., dissolved oxygen, temperature, salinity, pH) taken at discrete locations were averaged over the photic depth to also provide a composite value. This resulted in 42 variables representing the physical, chemical, and biological conditions in the reservoir within the photic depth. Each of the variables was also checked for normality distribution, correlations, and was transformed for ANFIS use. The reservoir's water column stability was examined in order to estimate its role in nutrient mixing, cyanobacterial density, and possible influence on metabolites concentrations in the reservoir (Journey, et al., 2008). Since density stratification (caused by thermal differences) was considered a possible determinant of algal bloom-metabolite concentration interrelations, several approaches to represent stability of the reservoir were investigated – Relative Thermal Resistance to Mixing (RTRM), Richardson number (R_i), diffusion coefficient (Munk, Anderson, 1948)

and precipitation moving average. The data matrix for ANFIS models use was finally transformed using logarithmic transformation, before it was used for building the fuzzy logic models.

ANFIS Model – Correlation Test

A normality test using Shapiro-Wilk test (Table 4) indicated that the original dataset consisted of many variables that were not normaly distributed. Hence, Spearman's Rho, using a comprehensive system for analyzing data sets, Statistical Package for Social Sciences 17.0 SPSS (Cronk, 2002) was chosen for correlation between variables (Table 5). Correlations greater than +/-0.7 were considered strong, less than +/-0.3 were considered weak and correlations between +/-0.3 and +/-0.7 were considered moderate. Most of the variables were statistically significant (at a significance level of $\alpha = 0.01$). Twenty four (24) variables were found to be correlated to *Pseudanabaena spp.*, sixteen (16) were found to be correlated to *Planktothrix agardhii*, fifteen (15) were found to be correlated to geosmin and twenty two (22) were found to be correlated to 2-MIB. *Pseudanabaena spp*. had strong correlation with cyanophytes and diatoms (coefficient > 0.7), moderate with specific conductance, total dissolved solids, salinity, pH, alkalinity, total phosphorus, NO_3^-, total N, total hardness and Ca^{2+}, K^+. The remaining variables with weak (coefficient < 0.3) correlations included *Cylindrospermopsis raciborski*, chlorophyll a, photic depth, temperature, SO_4^{2-}, orthophosphorus, NH3-N and precipitation (moving average). *Pseudanabaena spp*. had moderate negative correlations with total dissolved solids, salinity, alkalinity, nitrate, total N and Ca^{2+}, and week negative correlations with photic depth, sulfate, conductivity, NH3-N, chlorophyll-a, Mg^{2+}, and precipitation. *Planktothrix agardhii* was positively correlated with chlorophyll a, geosmin, *Cylindrospermopsis raciborski*, specific conductance, total dissolved solids, salinity, TKN, 2-MIB, Cl^-, SO_4^{2-}, and Mg^{2+}. *Planktothrix agardhii* had moderate correlations with chlorophyll a, microcystin, Cl^-, SO_4^{2-}, Mg^{2+}, Na^+, and geosmin, and weak correlation with 2-MIB. Geosmin had moderate to weak correlation with *Planktothrix agardhii*, microcystin, temperature, SO_4^{2-}, total dissolved solids, specific conductance, salinity, TKN, 2-MIB, *Cylindrospermopsis raciborski*, selected cations as Mg^{2+}, Na^+, K^+ and diatoms. The metabolite 2-MIB had moderate correlation with

microcystin, salinity, total dissolved solids, specific conductance, temperature, SO_4^{2-}, total phosphorus, and Cl^-, and weak correlation with *Planktothrix agardhii*, photic depth, alkalinity, NH3-N, geosmin, total hardness, sum RTRM, sum R_i, average diffusion and selected cations (Mg^{2+}, K^+, Na^-). These correlation results were used to select the most relevant variables for developing ANFIS models. The cations were eventually excluded from modeling as they cannot be predicted in EFDC.

ANFIS Model – Relative Thermal Resistance to Mixing (RTRM)

RTRM accounts for the nonlinearity of density and is a good indicator of water column stratification stability (Kortmann et al., 1982). This stability influences the vertical transport of nutrients and dictates the oxygen concentrations in the hypolimnion as a strongly stratified water column may lead to anoxic conditions at the sediment /water interface. This anoxia may trigger nutrient and metal fluxes from the sediment. Also, analysis of water column stratification can be helpful in estimating percentages of external and internal nutrient loadings to reservoirs and the upward/downward direction of nutrient migration. Further, density stratification may limit cyanobacteria mobility and restrict access to nutrients that may be available at lower levels in the water column. The stratification dynamics combined with meteorological mixing events prompt nutrient dependent algal dynamics. Furthermore, as indicated by Taylor et al. (2006), reservoir mixing conditions are very important in determining not only distribution, growth, and decay but also concentrations of taste and odor compounds. The strong stratification in reservoirs leads to differences in concentration of taste and odor compounds within the water column, as indicated by Westerhoff et al. (2002, 2005).

The RTRM values in ECR were calculated using the ratio of density differences between two neighboring layers to the density difference between water at 5 ^{0}C and 4 ^{0}C (Kortmann, et al., 1982).

$$RTRM = \frac{Upper\ layer\ density - Lower\ layer\ density}{Density\ at\ 5^0C - Density\ at\ 4^0C} \quad (1)$$

Sum of RTRM values across the vertical profiles at each station was used in ANFIS models to represent general stratification/mixing trends in Eagle Creek reservoir during 2008-2010.

ANFIS Model – Richardson Number (R_i)/Diffusion Coefficient

The use of RTRM values for estimating mixing conditions were also compared with Richardson number (R_i) values for the whole water column analysis (i.e., sum of R_i). Richardson number is an attractive approach since it not only incorporates the effect of temperature-dependent density differences, but also the influence of wind force on the stratification in the water column. Richardson number was calculated using the relationship between buoyancy and shear forces (Chapra, 1997).

$$R_i = \frac{-\left(\frac{g}{q}\right)\left(\frac{\delta q}{\delta z}\right)}{w_\theta/(z_s - z)^2}$$

(2)

where, g – gravity, q- density w_Θ – wind shear velocity, z_s- the water surface elevation.

The water density in the Equation 2 was calculated using the water temperature measurements for reservoirs vertical profiles from monitoring stations provided by CEES. Water surface elevation data was obtained from USGS 03353460 Eagle Creek at Clermont station.

When the R_i is significantly greater than ~ 0.25 (critical level), a stable condition is assumed and when it is below the critical level value the turbulence is generated (Chapra, 1997).

Next, Richardson number was related to diffusion coefficient:

$$E(z) = \frac{E_0}{(1+\alpha R_i)^{3/2}}$$

(3)

where, E_0 – the diffusion coefficient at neutral stability, R_i – Richardson number, α-a numerical constant for stability function f(R_i) defined as 3.33 (Munk, Anderson, 1948). Nakamura and Hayakawa (1991) list references to other numerical expressions for this constant existing in the literature and based on observations and empirical determinations.

The calculated range of depth-averaged diffusion coefficients for Eagle Creek Reservoir in the period of three years varied between 9.23E-03 m^2/s and 2.4E-0.7 m^2/s with higher values corresponding to spring and fall turnover and lower during warmer summer months.

4.3. Hydrodynamic and Water Quality Modeling

EFDC is an open source, second order accurate model, which can be used for one dimensional (1-D), two dimensional (2-D), and three dimensional (3-D) simulations of rivers, lakes, reservoirs, estuaries, coastal areas and wetlands. EFDC has many worldwide applications and some of them were reviewed in this research to determine the accuracy and suitability of the modeling tool. Existing cases that were examined included but were not limited to studies of Lake Lanier (Tetra Tech. Inc., 2009), Lower St. John River (Tillman et al., 2004)' and Mashapang Pond (U.S. EPA, New England Region, 2002). In all cases, EFDC was shown to achieve satisfying results.

EFDC has hydrodynamic, sediment-toxic fate and transport and water quality components. The hydrodynamic component solves depth averaged Navier–Stokes equations assuming incompressible flow and hydrostatic pressure distribution with dynamically coupled salinity and temperature transport. The EFDC hydrodynamic module also incorporates wetting/ drying of cells, vegetation resistance, and control flow structures options, which allows more realistic simulations of natural and engineered conditions (Tetra-Tech., 2007).

Equations 4 and 5 represent the change in momentum as a result of pressure force, eddy viscosity, and sources/sinks.

$$\frac{\partial(Hv)}{\partial t}+\frac{\partial(Huv)}{\partial x}+\frac{\partial(Hvv)}{\partial y}+\frac{\partial(vw)}{\partial z}-fHu=-H\frac{\partial(p+g\eta)}{\partial y}+\left(-\frac{\partial h}{\partial y}+z\frac{\partial H}{\partial y}\right).\frac{\partial p}{\partial z}+\frac{\partial}{\partial z}.\left(\frac{A_V.\partial v}{H\partial z}\right)+Q_V$$

(4)

$$\frac{\partial(Hu)}{\partial t}+\frac{\partial(Huu)}{\partial x}+\frac{\partial(Huv)}{\partial y}+\frac{\partial(uw)}{\partial z}-fHv=-H\frac{\partial(p+g\eta)}{\partial x}+\left(-\frac{\partial h}{\partial x}+z\frac{\partial H}{\partial x}\right).\frac{\partial p}{\partial z}+\frac{\partial}{\partial z}.\left(\frac{A_v.\partial u}{H\partial z}\right)+Q_u$$

(5)

Where,

H - water column depth

u, v, w - velocity components in the curvilinear, sigma, x-, y-, and z- directions, respectively

A_v - vertical turbulent momentum mixing coefficient

Q_v , Q_v - internal and external sources and sinks

m_x, m_y - horizontal curvilinear coordinate scale factors

Another basic governing equation for the hydrodynamic module of EFDC is conservation of mass (Equation 6).

$$\frac{\partial H}{\partial t}+\frac{\partial(H_u)}{\partial x}+\frac{\partial(H_v)}{\partial y}+\frac{\partial w}{\partial z}=Q_H$$

(6)

The transport equations for salinity and temperature (Equation 7 and Equation 8) respectively are solved in the hydrodynamic module, where the source and sink terms Q_s and Q_T include subgrid scale horizontal diffusion and thermal sources and sinks, and where A_b is the vertical turbulent diffusivity (Hamrick, 1992).

$$\partial_t(mHS)+\partial_x(m_yHuS)+\partial_y(m_xHvS)+\partial_z(mwS)=\partial_z(mH^{-1}A_b\partial_zS)+Q_s$$

(7)

$$\partial_t(mHT)+\partial_x(m_yHuT)+\partial_y(m_xHvT)+\partial_z(mwT)=\partial_z(mH^{-1}A_b\partial_zT)+Q_T$$

(8)

The water quality component, linked internally with hydrodynamics, consists of an eutrophication module based on 21 variables including three functional algal groups, refractory, labile and dissolved fractions of carbon, phosphorus, nitrogen, total phosphorus, ammonium, nitrate and nitrite, silica, dissolved oxygen, chemical oxygen demand, total suspended solids, total active metal, and coliform bacteria. The module simulates nutrient fluctuations in the water body and associated fluctuations in algal communities. The eutrophication model solves mass balance equations (Equation 9) for all of the 21 variables:

$$\frac{\partial C}{\partial t} + \frac{\partial (uC)}{\partial x} + \frac{\partial (vC)}{\partial y} + \frac{\partial (wC)}{\partial z} = \frac{\partial}{\partial x}\left(K_x \frac{\partial C}{\partial x}\right) + \frac{\partial}{\partial y}\left(K_y \frac{\partial C}{\partial y}\right) + \frac{\partial}{\partial z}\left(K_z \frac{\partial C}{\partial z}\right) + S_c$$

(9)

Where,

C – Concentration of state variable;

u, v, w – velocity components in x, y, z directions;

K – Turbulent diffusivities in the x, y, z directions;

Sc – Sources and sinks per unit volume.

Equation 9 consists of physical transport and kinetic processes terms. The last three terms on the left-hand side account for advective transport, and the first three terms on the right-hand side account for diffusive transport. The last term on the right-hand side represents kinetic processes and external loads for each of the state variables (Hamrick, 1992). The model solves Equation 9 after decoupling the kinetic terms from physical transport as they have different time scales. The decoupling of mass balance in Equation 9 makes the model more flexible allowing additional state variables. The kinetic processes are formulated in Equation 10:

$$\frac{\partial C}{\partial t} = [K] \cdot [C] + [R]$$

(10)

K – kinetic rate (time^{-1});

C – concentration (mass volume$^{-1)}$;

R – source/sink (mass volume^{-1} time^{-1})

In this study, EFDC hydrodynamic and water quality modules were applied to simulate the transport and fate of three algal groups (blue-green algae, green algae and diatoms), and main water quality constituents.

4.4. Adaptive Network-based Fuzzy Inference System (ANFIS)

Neuro-fuzzy modeling using Adaptive Network–based Fuzzy Inference System was selected for prediction of odor and taste compound release events accompanied by cyanobacterial algal blooms. First introduced by Jang (1993), ANFIS is designed to deal with uncertain dynamic systems often supported by sparse, ill-defined data. ANFIS is a simple data learning technique that uses fuzzy inference system (FIS) model based on concepts of fuzzy set theory and fuzzy if-then rules (Ghani, 2009). The ANFIS modeling encompasses fuzzy reasoning and the adaptive networks. In this research, Takagi-Sugeno-Kang fuzzy system developed in 1985 was used (Takagi, Sugeno, 1985). The main advantage of using ANFIS is its fast learning and adaptation capabilities and the fact that after the proper training, ANFIS can bypass use of iterative processes for new introduced cases (Turkmen et al., 2009). In ANFIS, the fuzzy inference system is generated by a hybrid learning process, which incorporates both the back propagation gradient descent and least squares methods. The structure of ANFIS is composed of 5 layers (Figure 12). Here, for simplicity of explanations the two inputs (x, y) system is presented (ANFIS supports multiple input, single output systems).

In Figure 12, square nodes (adaptive nodes) demonstrate that the parameters in these nodes are adjustable, to be learned, while the circle nodes (fixed nodes) demonstrate they are fixed parameters. A common rule set with two fuzzy if-then rules is as follows:

Rule 1: If x is A_1 and y is B_1, then $f_1 = p_1 x + q_1 x + r_1$

Rule 2: If x is A_2 and y is B_2, then $f_2 = p_2 x + q_2 x + r_2$

Where A, B are linguistic terms that are user defined and representing a range of values.

The sequence and functions of the layers is as follows:

Layer 1: Square node equipped with node function

$$O_i^l = \mu_{A_i}(x),$$

(11)

Where, O_i^1 is the membership function of A_i and x is the input parameter to the node. The A_i is the linguistic label connected with the node function.

Layer 2: Circle node described in Figure 10 as $\prod$. This node multiplies the incoming signal and sends the product out. Each node output is the firing strength of a rule.

$$w_i = \mu_{A_i}(x) \times \mu_{B_i}(y), \ i = 1,2$$

(12)

Layer 3: Circle node. Node calculates the ratio of the i-th rule's firing strength to the sum of all rule's firing strengths:

$$\overline{w_i} = \frac{w_i}{w_1 + w_2}, i = 1,2$$

(13)

Layer 4: Square node with node function

$$O_i^4 = \overline{w_i} f_i = \overline{w_i}(p_i x + q_i y + r_i)$$

(14)

p, q, r – parameter set (consequent, linear, parameters)

Layer 5: Circle node. This node computes the overall output as summation of all incoming signals.

$$O_i^5 = overall\ output = \sum_i \overline{w_i} f_i = \frac{\sum_i w_i f_i}{\sum_i w_i}$$

(15)

Hybrid learning algorithm uses two passes, forward and backward. In the forward pass, the consequent parameters are identified by the least square estimate and in the backward pass, the error rates propagate backward and the premise parameters in layer 1 described in Gaussian membership function μ_{Ai} (nonlinear parameters to be learned) are updated by gradient descent (Jang, 1993). The overall output is represented by polynomial - f_i.

4.5. Model Development

4.5.1. EFDC Hydrodynamic Model

The hydrodynamic and transport model was configured for Eagle Creek Reservoir for 2008. The first 219 days were used for calibration and the last146 for validation.

Grid Generation

The grid was generated using EFDC grid generating capabilities. In order to find the best fitting resolution of grid cells for reservoir, several variants have been investigated (Table 6). Since EFDC allows only for single focal point with expanding Cartesian grid, three expanding and one uniform grid system were generated with the focal point at the water intake to better simulate the complex flow around the gates and near the causeway. Zero cell rotation angle defined the causeway under the bridge in a satisfying manner and was chosen in the final grid. The grid size was carefully chosen to meet the requirements defined in Courant- Friedrichs- Lewy (CFL) condition, which is expressed as:

$$\gamma = U\frac{\Delta t}{\Delta x} \quad (16)$$

Where, γ- Currant number, U-velocity, Δt-time step, Δx-cell size

The restriction $\gamma < 1$ for grid size and time step ensures numerical convergence and stability.

The final cell size with expanding factor of 1.005 varied from 40 to 60 m and had zero cell rotation. The reservoir grid composed of 2401 cells and five vertical layers of varying depth.

Initial/Boundary Conditions

Initial conditions for EFDC hydrodynamic module include:

a) Initial water surface elevation
b) Water temperature
c) Salinity

Initial water surface elevation was 240.56 m according to USGS measurements on January 1st, 2008. Water temperature of 3.4 ^{0}C at Mill Creek USGS gage station near Manhattan, IN on January 1, 2008 was used for the initial condition. This value of water temperature recorded in a different ecoregion was used due to insufficient data for the Eagle Creek Watershed region and after comparison analysis of water temperature between Mill Creek USGS gage station # 03359000 and USGS gage # 03354000 at White River near Centerton located 25 miles from ECR. The coefficient of determination (R^2) for the regression was 0.90. The water temperature value for initial salinity was chosen as 0.25 kg/m^3, based on the average value for ECR in 2008.

The hydrodynamic boundary conditions of Eagle Creek Reservoir included estimations of time-series of inflow discharge from 12 tributaries (via the SWAT model), measured outflow discharges at the water intake and ECR dam, and atmospheric conditions including wind speed, wind direction, precipitation, evaporation, cloud cover, and solar radiation. The tributary inflows from sub-basins and outflows from the reservoir were assigned grid cells corresponding to the physical location in the reservoir (Table 7, Figure 13).

4.5.2. Water Quality Model

The hydrodynamic eutrophication model (HEM-3D) was simulated for 19 constituents (3 algal functional groups, labile, refractory and dissolved fractions of phosphorus, nitrogen and carbon, total phosphorus silica, dissolved oxygen and chemical oxygen demand) for 2008, with 285 days serving as the calibration period. Initial conditions for water quality variables were established based on the annual averaged nutrient values for Eagle Creek Reservoir on January 1, 2008. The initial conditions applied to the model consisted of partitioned (dissolved, refractory, labile) nutrients, and were considered spatially constant for the whole study area. Additionally, the model inputs included the literature based constants comprised of values for algal dynamics, stoichiometry, light extinction, half-saturation values for algae and nutrients, and temperature (Cerco, 2000; McGovern, 2006; Ernst and Owens, 2009; Tetra Tech, 2009). These constants were later adjusted during the calibration process as they can vary depending on site-specific conditions.

The boundary conditions for the reservoir were based on SWAT - based predictions of time-series of nutrient concentrations in the 12 tributaries. These nutrients were partitioned into the dissolved, refractory and labile fractions of organic phosphorus and nitrogen. The nutrients also incorporated total phosphorus, inorganic nitrogen, dissolved oxygen, ammonium and available/unavailable silica. Carbon concentrations were partitioned into dissolved, labile and refractory fractions, based on watershed averages, and uniformly applied to the reservoir's tributaries.

5. Results and Discussion

5.1. Data – Conditions in the reservoir

The climatic data over the three year period showed a 1 ^{0}C increases in average air temperature in 2009 and 2010 in comparison with 2008, and also decreased precipitation in spring and summer seasons (Table 8, Figure 14). In general, each of the three years (2008-2010) showed similar seasonal trends in algal dynamics and metabolite concentrations, but varied in terms of intensity of algal abundance and by-product concentrations from year to year. The observed trends in ECR demonstrated increased cyanobacteria productivity during late summer/early fall seasons 2008 and 2009 and two biovolume peaks in 2010: spring and late summer. It has been noted that sampled cyanobacterial maximum biovolumes increased over the 2009 and 2010 relative to 2008 data (Figures 15-17) and the highest net growth rates were observed in July of 2010 (0.8 d^{-1}). The estimates of net growth rates were based on assumption of no nutrient limitation as presented in Wong et al. (2009). It was noted that the net growth rate was positively correlated with temperature, and relative resistance to mixing (RTRM), but also had negative correlation with total phosphorus concentrations and total dissolved solids. There was no spatial growth rate variation between the main sampling regions (marina, dam, intake); however, strong seasonal dependence was observed with maximum growth rates in the months of July and August (Table 9, Figure 18). The maximum growth rates for cyanobacteria observed in late summer and early fall are explained by cyanobacteria's preference for high water temperatures (Domingues et al., 2011). The growth rates were further adjusted in EFDC software based on actual water quality dynamics.

The two T&O producing cyanobacteria species, *Planktrothrix agardhii* and *Pseudanabaena spp*. present at ECR that were investigated in this study varied over the course of three years 2008 -2010 and showed clear seasonal distributions during 2008-2010 (Table 10). *Pseudanabaena spp.* was the most abundant in late summer/early fall during these three years and its highest biovolume was observed in the late summer of 2010 (08/10/2010) (Figure 19). *Planktrothrix agardhii* was the most abundant in 2009 and reached its highest biovolumes on 10/15/2009 (Figure 20). In comparison to *Pseudanabaena agardhii, Planktothrix agardhii* was represented by significantly smaller biovolumes during 2008-2010 but both of the species bloomed during the same time.

Reservoir water column stability was examined because of the widely reported linkages between reservoir stability and nutrient availability, cyanobacterial density, and the possible influence on metabolites production in reservoirs (Journey et al., 2008; Westerhoff et al., 2002; Taylor et al., 2006, personal communication with Dr. Tedesco and Clercin N.). Several approaches to represent stability of the reservoir were investigated: Relative Resistance to Thermal Mixing (RTRM), Richardson number (Ri), diffusion coefficient, and precipitation moving average.

Sum of RTRM for each vertical station profile was used to describe trends in Eagle Creek reservoir during 2008-2010, where the value 30 for sum RTRM was used to define the changing reservoir conditions (Tedesco, personal communication). The stability calculations were based on bi-weekly observations at the monitoring stations, which limit the interpretation of the water column conditions between sample collection events.

In general, the 2008 data for near the dam location showed that the water column was stratified during sampling events until September 30, 2008. Similarly, the marina region was stratified for most of the late spring/summer fall season until September 3, 2008, when it became mixed. The mixing condition observed on June 05, 2008 was correlated to increased precipitation during the preceding days. In 2009, dam stations showed increased stratification during early May sampling events (May 4, 2009) and remained stratified until October 1. The intake stations were well mixed before May, but

experienced stratification until the October 15 sampling event, after which the water column became well-mixed again. In 2009, the shallower marina stations experienced fluctuations of water column stratification. Although, mostly stratified, it experienced a few mixing events during the summer (07/09/2009, 08/16/09) and then completely mixed September 17, 2009. In 2010, the reservoir near the dam remained stratified with the first measurement indicating increased thermal resistance to mixing on April 21. The system was mixed on September 30, 2010. The intake stations were stratified throughout the monitoring season (spring, summer, and fall) until September 8, 2010 with the exception of May 18. The water column at the shallower marina stations was stratified from April 4, 2010 to September 21, 2010, when it mixed. On October 15, water column stratification increased again. The 2010 water conditions were linked to observed increased temperatures and decreased precipitation during the summer season (low pool elevation), which also coincided with the increased levels of metabolites observed in the reservoir.

These calculations of RTRM were cross examined with Richardson number calculations, which similar to RTRM were based on whole water column analysis (i.e. by using SUM of R_i for all vertical depths at a X-Y location) and agreed with general trends of water column stability (Figures 21-28). Although, it was acknowledged that the spatial mixing may exists within the water column at a given time due to local mixing factors, the sum of R_i, was chosen as a general representation of the stability conditions in the reservoir. The results shown in the Figures 21-28 clearly showed the existing mixing conditions during spring and late fall seasons and the R_i value reflected the atmospheric influences over the summer months, which were effecting stability of the water column on a smaller scale. The discrepancies between sum RTRM trends and sum R_i existed for several dates in 2009 and 2010, where the decreased value of Richardson number corresponded to increased value of the sum RTRM. The fifteen day precipitation moving average was used to explain this difference. The dates of 5/19/2009, 7/1/2009, 7/2/2009 and 7/28/2010 corresponded to values of 15 day precipitation moving average of 6.08 mm, 2.20 mm, 2.24 and 8.75 mm, respectively. This atmospheric influence on water column was captured by R_i, but was not shown in the RTRM results (Appendix, Figures I-IV). The

9/17/2009 precipitation value was insignificant (0.20 mm) and can not be used to explain decreased Richardson number value and RTRM variation. The decreased value of RTRM was more likely caused by the seasonal temperature change.

Similarly, the diffusion coefficient at neutral stability (m^2/sec) and the diffusion coefficient averaged over the water column for each of the stations were calculated for the period of three years 2008-2010. The results showed that in 2008 and 2010 the reservoir water column was stable, which was represented by small values of diffusion coefficient. The most variation was observed in 2009, where the calculated averaged diffusion coefficient went over the diffusion coefficient at neutral stability during the times corresponding to precipitation. The precipitation on 5/19/2009 was not reflected in diffusion coefficient results, which have shown stabile conditions in the ECR in spite of precipitation event.

All these calculations of the water column stability were developed to help in decision making of the most appropriate parameters for fuzzy models development and also to check their agreement. In the final parameter selection for representing the mixing conditions in ECR, the data based R_i calculations were further used in fuzzy models as its value is also spatially and temporally predicted in EFDC software.

The statistical analysis of geosmin and 2-MIB for period 2008-2010 showed increasing concentrations of these metabolites within the reservoir, which exceeded the level of human detection, which has the literature described values of 5 ng/L (Dzialowskiet al., 2007) to 10ng/L (Hobson et al., 2010). The maximum recorded geosmin and 2-MIB concentrations during the three years occurred in 2010 with geosmin of 109. 43 ng/l and 2-MIB of 223.72 ng/l (Table 11, Figures 29-34).

5.2. Hydrodynamic Model Calibration/Validation

The calibration/validation of the hydrodynamic model involved surface elevation and temperature adjustments based on observed and simulated values. The model was calibrated/validated to water surface elevation recorded at USGS gage station #

03353450 in the Eagle Creek Reservoir and to temperature profiles collected during 2008. The model was calibrated for 219 days (January 1, 2008 - August 7, 2008) and validated for 146 days (August 7, 2008 – December 31, 2008). The surface elevation calibration was achieved by calculation of absolute error between the observed and simulated total water volume and the adjustment to the tributary inflows. An average RMSE of 0.091 m and 0.023 % relative error were attained from 218 data pairs at the end of the surface water elevation calibration process (Figure 35). The coefficient of determination (R^2) for the calibration period was 0.86 (Figure 36). The validation result for average RMSE was 0.102 m and 0.022 % relative error for 153 data pairs (Figure 37). The R^2 for the validation period was equal to 0.94 (Figure 38). The model predicted surface elevation quite well and the relative error stayed within the error of the Raytheon Echo Sounder–Model 719 C used for bathymetry measurements.

The EFDC initial simulations were based on solar radiation recorded at TPAC station and were compared to results from simulations with solar radiation internally computed during the water temperature calibration process. The model results for TPAC solar radiation were more accurate and used in the final calibration. Water temperature calibration included 23 monitoring stations (ECR Stations: A-F) totaling 181 data pairs. Calibration was finalized with average RMSE of 1.28 ^{0}C and relative error of 5.26 % (Table 12, Figures 39 a-u). The validation RMSE was 1.62 ^{0}C and relative error 6.61 % for 31 monitoring stations (ECR Stations: G-N) and 235 data pairs (Table 13, Figures 40 a-zc). The model had an overall reasonable representation of temperature observations however it generally did not predict stratified conditions accurately.

5.3. Water Quality Model Calibration/Validation

Water quality calibration was done for 285 days and was performed after hydrodynamic model calibration and validation. The period of time chosen for calibration was to capture the differences in algal dynamics during the spring, summer and fall of 2008. The model was calibrated for cyanobacteria, diatoms, green algae, total phosphorus, nitrate-nitrogen, dissolved oxygen and ammonia-nitrogen. The model predicted values for chosen calibration parameters were statistically compared to observed concentrations at

the monitoring stations. The statistical analysis included relative and root mean square error computations. The calibration was performed for 3 representative monitoring stations located in the dam, and marina region. There were a total of 44 water quality runs. Model calibration involved manual adjustments of a subset of several parameters chosen from the total of 300 model parameters. The sensitive parameters used in the final calibration included: algal maximum growth rates, basal metabolism for three algal groups, nitrification rates, phosphorus and nitrogen half-saturation values for algae functional groups, predation on algal groups, reaeration rates, minimum dissolution rates of refractory, labile organic carbon and lower and upper optimal temperatures for algal growth.

The RMS error for total phosphorus for dam monitoring station varied from 0.058-0.399 mg/L (27 data points). The overall average RMS error for the total phosphorus for this location was 0.191 mg/L (Figures 41 a-b). The RMS error for marina location (25 data points) fluctuated between 0.066 and 0.952 mg/L, with the average RMSE of 0.288 mg/L. The model significantly under-predicted total phosphorus. This was attributed to estimated tributary inputs and a lack of sediment flux influences. While the field measurements captured the increased concentrations of total phosphorus near the reservoir bottom, the model's scale of increase was significantly smaller.

The RMSE for ammonia-nitrogen for dam monitoring station (27 data points) varied between 0.021 to 2.399 mg/L with the average RMSE of 0.957 mg/L. The marina station ammonia-nitrogen calibrated ended with the resulting average RMSE of 0.065 mg/L, where the error varied between 0.019 mg/L and 0.111 mg/L. Ammonia-nitrogen was under-predicted (Figures 42 a-b). Just as in the phosphorus case the variation in error was linked to the unknown sediment fluxes not simulated in the model and also to the estimation of the watershed inputs.

The nitrate-nitrogen was over-predicted by the model and RMSE varied between 0.613 mg/L to 2.063 mg/L for dam station (27 data points) and 0.280 mg/L to 5.74 mg/L for marina (28 data points) with the average RMSE of 1.78 mg/L and 2.77 mg/L,

respectively (Figures 43 a-b). The higher end of the error statistics represented the immediate response to the discrepancies between the actual nutrient inflow into the reservoir and values estimated in the SWAT model.

Dissolved oxygen, crucial in many of the biological, chemical and physical processes for which it is a main driver, had an average RMSE of 6.99 mg/L for dam monitoring station (150 data points), 5.45 for intake monitoring station (97 data points) and 4.61 mg/L for marina station (62 data points). The RMSE for dissolved oxygen varied between 1.84 to 18.01 mg/L (Figures 44 a-c). The oxygen inputs into the model had an impact on the overall model performance and the watershed loads would have to be reexamined to sufficiently judge the quality of the calibration. It has been observed that prediction of dissolved oxygen was most of the times uniform in the vertical water column profile, and in a few cases the increase of oxygen concentration was observed in the middle and the bottom of the reservoir. This indicates that the oxygen demand (or production by cyanobacteria blooms at depth) was not well represented by the model. The substantial RMSE reflected the high oxygen concentrations sporadically predicted by SWAT model. The simulated algal groups included cyanobacteria, diatoms and green algae. The cyanobacteria for dam monitoring station were predicted with the resultant RMSE of 0.514 mg C/L. The averaged observed concentration for this station was 0.605 mg C/L while the modeled averaged value 0.22 mg C/L (Figure 45). The presented minimum, average and maximum curves shown in the Figure 45 were internally calculated within the EFDC software. The minimum, average and maximum option for presenting the output generated three model time series for the cell, based on the water column layer results.

Diatoms calibration was finalized for dam monitoring station with RMSE of 0.302 mg C/L. The average observed value was 0.23 mg C/L and the average modeled concentration 0.176 mg C/L (Figure 46). The green algae were not calibrated as the modeled values were in a range of E-04 mg C/L while average observed value was 0.399 (Figure 47).

The average RMSE for cyanobacteria for the marina region was 1.28 mg C/L, where the average observed value for cyanobacteria was 1.29 mg C/L and the average modeled 0.38 mg C/L (Figure 48). The cyanobacteria was under predicted in the first part of the simulation (May-August) and over predicted in September, 2008 (Table 14). The diatoms for the marina monitoring station were also under predicted with the final RMSE of 0.258 mg C/L. The averaged observed value for diatoms was 0.22 mg C/L and the averaged modeled concentration of 0.108 mg C/L (Figure 49).

The green algae failed to be predicted with the average RMSE of 0.500, where the average observed concentration was 0.432 mg C/L and the modeled value was in order of E-04 mg C/L (Figure 50). The green algae was underpredicted and was eventually removed from calibration as it was not selected as one of the variables for ANFIS model. The statistics for algal calibration results are presented in Table 14 and Table 15.

In comparison with cyanobacteria, diatoms were over predicted during first few months of the model simulation (May-July) and later under predicted.

The water quality model requires further calibration based on the obtained results as soon as the watershed inputs are reexamined and improved. The boundary conditions for a model are very important for achieving the required accuracy. Here, the model was accepted for further application with the awareness that the results may affect the overall output of the developed methodology for spatial and temporal prediction of metabolites. To effectively improve model output, it would be necessary to obtain calibrated SWAT model results to establish the correct boundary conditions, include the sediment fluxes in the water quality modeling as well as reexamine to calibration parameters. The manual calibration is a time consuming process, while working with the multiple water quality parameters.

5.4. ANFIS Model Result

This section describes the development of the ANFIS models for Eagle Creek Reservoir based on data analysis results. The input/explanatory variables selected for these models

consisted of variables that had acceptable correlations with the response or output variables in these models, and were variables that were also simulated by EFDC-HEM3D models. Four models were developed, each of which predicted one of these response variables - *Pseudanabaena spp*., *Planktothrix agardhii*, geosmin and 2-MIB.

5.4.1. ANFIS 1 - *Pseudanabaena spp.* Model (1)

The data matrix used for building the *Pseudanabaena spp.* fuzzy logic model included selected variables that had (a) reasonable to significant positive /negative correlations with *Pseudanabaena spp.*, and (b) could be predicted using the EFDC model. Additionally, the data matrix also included easily acquired atmospheric variables. The matrix contained: cyanophyte and diatom biovolumes, temperature, salinity, total phosphorus, nitrate and precipitation moving average. Variable matrix was log transformed before it was trained in Matlab using ANFIS inference. There were only 112 data points available for training and checking for seven variables, so 102 data points were used for training and 10 data pairs for checking model performance. ANFIS performance was constrained by this proportion of variables and available data. The initial FIS was generated using sub-clustering with radius r=0.5. Three membership functions (MFS) were created and new FIS was optimized using hybrid method. Error tolerance was set to 0.001 and training was completed after 2 epochs with the training error (RMSE) of 0.33 mm^3/m^3. Training RRMSE of 65% the trend lineR^2 of 0.70 and coefficient of determination of R^2 equal 0.70 were found to be satisfactory for further model validation (Figures 51-52). The evaluated FIS was applied to validation data set (Table 16, Figures 53-54) The RRMSE for validation dataset was 45 %, and R^2 was equal to 0.72 (R^2 = 0.79). The validated model's fuzzy inference structure (FIS) was next applied to grid cells of the EFDC model to create the spatial temporal distribution of *Pseudanabaena spp*. in the reservoir. The available monitoring stations served as model's performance guidance.

5.4.2. ANFIS 1 - *Planktothrix agardhii* Model (2)

Available data for used in *Planktothrix agardhii's* model was divided in the same manner as the *Pseudanabaena spp.* model, which included 102 data pairs for training the

model and 10 data pairs for testing. *Planktothrix agardhii* had weaker correlations with environmental variables in comparison with *Pseudanabaena spp*. The strongest correlations were with microcystin, SO_4^{2-}, Cl^-, Mg^{2+}, Na^- and geosmin. None of these variables were incorporated in the modeling scheme as they are not predicted by EFDC module. The model's matrix was initially composed of weakly correlated variables including diatoms, chlorophyll-a, salinity, NO_3 and TKN. Model simulations produced large training error (RMSE) > $1mm^3/m^3$. To improve model performance TKN was removed from matrix (weak correlation: 0.195; significant at 0.05 level) and replaced by total phosphorus. This decision was based on modeler's knowledge about *Planktorthix agardhii's* moderate affinity for phosphorus and also on existing literature review indicating phosphorus as important variable for determining the specie's biomass abundance (Ducobu et al., 1998; Arnaud et al., 2008; Aubriot et al., 2011). The FIS was created using subtractive clustering with radius of 0.5. As the result of sub-clustering, the six Gaussian membership functions were created and a hybrid optimization method was used for model fitting. The FIS was trained with average training error (RMSE) of 0.68 mm^3/m^3, RRMSE of 165 % and coefficient of determination $R^2 = 0.11$ and excel plotted $R^2 = 0.56$ (Figures 55-56). In general, the generated FIS overpredicted the species biovolumes in 2008 and missed to capture the high extremes. The model underpredicted the high biovolumes and overpredicted low for the remaining two years (2009-2010). Although not precise, the model was considered for further testing and tried on the unseen by model data. The model under-predicted the highest observed biovolumes and did poorly for this high range, while low values were slightly overpredicted. The coefficient of determination R^2 for validation was 0.86 and RRMSE 86 %. The validation data set was under-predicted for high observed values of biovolumes and overpredicted for the low observed values of biovolumes (Table 17, Figure 57). In general model performance was considered poor and relatively good result in the validation data set possibly considered coincidental.

5.4.3. ANFIS 2 - Geosmin Model (1)

Geosmin model consisted of 6 correlated input variables (*Pseudanabaena spp*., diatoms, *Planktrothrix agardhii*, water temperature, salinity and TKN) and one output variable -

geosmin. The variable selection was based on correlation results and modelers knowledge. *Pseudanabaena* spp. was incorporated based on modeler's knowledge as it has been a known geosmin producer. In comparison with a previous study (Dzialowski et al., 2007) the correlation results didn't confirm a strong relationship between total cyanobacteria biovolume, sechii disc depth and phosphorus, and thus were not chosen as strong geosmin predictors. Similarly to other ANFIS models, the data set was divided into 102/10 pairs for training and validating. The initial FIS was generated using subtractive clustering with radius 0.5 which produced four MFS. After the clusters have been determined, hybrid learning procedure with an error tolerance of 0.01 was used for estimation of the premise and consequent parameters. Training was finished after 2 epochs. The average training error (RMSE) was 0.17 ng/l, RRMSE = 55 % and $R^2 = 0.84$ (equal to excel plotted R^2) (Figures 58-59). The evaluation of observed versus predicted was accepted as satisfying. There was no general trend in prediction deviations observed as occasionally the model under-predicted the high values as well as over-predicted low values. Model was validated using the validation data matrix. The R^2 for validation data was 0.78 ($R^2 = 0.87$) and RRMSE 57 % (Table 18, Figure 60). Figure 60 shows that the predicted values of geosmin followed a general trend observed of increased and decreased geosmin concentrations in ECR. Significant peak concentrations were under-predicted (especially observation on 9/17/2009) by model, while low concentrations show a good fit between observed and predicted values. The model was considered as a good fit for this study.

5.4.4. ANFIS 2 - 2-MIB Model (2)

The input variables for the 2-MIB ANFIS model comprised of water temperature, total phosphorus (TP), *Planktothrix agardhii*, salinity and sum of Richardson number and *Pseudanabaena spp*. First five (5) variables were selected based on medium correlation with 2-MIB, and the last one was added since it is a known 2-MIB producer. Although, NO_2 was negatively correlated with 2-MIB (-0.301, at a significance level of 0.01), it was not included in the model due to its short life time in the water column and also is not predicted by EFDC. The FIS was generated using grid partition and bell membership function. Two membership functions were generated as part of the model fitting process.

The training error was 0.19 ng/l RMSE, R^2=0.82 and RRMSE=71% (Figures 61-62). The model exhibited a trend of over-predicting 2-MIB values in the lower range and under-predicting high concentrations, but maintained the general 2-MIB concentration trends. The fuzzy inference structure was next applied to the ten randomly selected data points. The two out of ten values were significantly large, which didn't allow to establish the coefficient of determination for this small validation set (Table 19, Figure 63). The overall performance of the model was found to be fair to poor.

5.5. ANFIS –EFDC Application for species and metabolite predictions

All four ANFIS models were trained and tested on field data. Once the EFDC water quality results were available, trained and tested FIS were first applied to the selected EFDC reservoir grid cell with known field measurements (L3: Easting-559341.0, Northing-4412667.0) for cyanobacteria species and metabolites observed-predicted values comparison. Next, the top layer (layer five) predictions from EFDC model were combined with ANFIS to create the prediction for the whole reservoir for a selected date. The selected grid cell had records for the the water quality parameters for September 30, 2008. The FIS evaluated grid cell was located in the north basin of ECR (Figure 64). The EFDC predictions for cyanobacteria and diatoms were converted from concentrations of mg Carbon/Liter to biovolumes (mm^3/m^3) using the average carbon per unit species (pg/cell) recorded for ECR. This transformation allowed on species cell estimation, which next was multiplied by unit species biovolume to obtain the final biovolume. This was done to convert the units of EFDC outputs into units of ANFIS model inputs, since the ANFIS was trained and tested on log transformed biovolume values.

The predicted output for selected grid cell for *Pseudanabaena spp*. was 18,386,540 mm^3/m^3, while the observed value 942,478mm^3/m^3. The ANFIS 1 model significantly over predicted this cyanobacteria species biovolume. The *Planktrothrix agardhii* prediction from ANFIS 1 model was a negative number, which was assumed to represent zero value. This corresponded to a count at selected location, where *Planktothrix agardhii* was not observed in September of 2008 (based on CEES reported data). These predicted values were then used as inputs to ANFIS 2 models for metabolite prediction.

The new simulation was run and the predicted geosmin value for selected location was 3.43 ng/l, while the observed value was 11.35 ng/l. The model underpredicted this metabolite concentration.

The 2-MIB prediction was not validated as the ANFIS output based on EFDC and ANFIS species results gave the value, which was outside the reasonable concentration range (E+20). Finally, the FIS was applied to variable outputs for 4 constructed models for 2401 EFDC grid cells. The *Pseudanabaena spp*., *Planktothrix agardhii* and geosmin prediction is shown in Figures 65, 66 and 67. The 2-MIB was not mapped due to it's out of range issue.

Based on the performed simulation, it was observed that on September 30, there was no *Planktothrix agardhii* present in the ECR.

The predicted *Pseudanabaena spp.* biovolumes were the highest in the central part of the ECR north basin in a range of 17,554,967 mm^3/m^3 – 20,422,081 mm^3/m^3. The predicted biovolumes decreased in north and south directions and reached the lowest values in the area of the Eagle Creek and School Branch tributary inflows.

Geosmin spatial distribution for September 30 showed the highest geosmin concentrations in the sparse locations on the east shore of the ECR north basin, south basin and in a single cell in the School Branch tributary inflow. There were more grid cells with the increased geosmin concentration with a range at human detection level of 4.92-9.74 ng/L located in the northern part of the ECR (east shore). Most of the reservoir fell in the medium range of predicted geosmin with concentrations of 2. 88 ng/L to 4.91 ng/L and the lowest values were predicted in the Eagle Creek and School Branch tributary inflows. The fact that 2-MIB and geosmin producing *Planktothrix agardhii* was predicted at zero value for the selected date and the fact that there were detected concentrations of geosmin at the human detection level didn't allow determining their spatial relationship between the producer and the metabolite. Also, *Pseudanabaena spp.* a known producer of 2-MIB predicted in the ANFIS model couldn't be compared with the

spatial distribution of 2-MIB as this metabolite wasn't validated due to large errors produced during the validation.

Conclusions

The results of this study indicate that the developed methodology incorporating the mechanistic and heuristic modeling could be applied in reservoir taste and odor algal metabolite predictions. However, there were several limitations which influenced the overall model performance and its capability to follow the observed trends in the reservoir. The main sources of error were the state of knowledge in terms of understanding the metabolite contributions from different sources, data availability and data accuracy. The ANFIS models were built without information regarding the taste and odor contributions from algal benthic species, soil bacteria and possible watershed inputs. If available, these data could further improve the models performance. The watershed metabolite input issue could be resolved within the proposed methodology based on the metabolite spatial prediction if other data were available. Although not validated, the ANFIS Geosmin model prediction showed that the highest concentrations of geosmin could be expected in the small areas along east shore in close approximidity to 56th Steet Bridge. The predicted elevated geosmin concentrations ranging from 4.92 ng/L to 9.74 ng/L are located in the shallow areas of the reservoir with depth less than 2.5 meters. The other shallow regions had concentrations less than 4.52 ng/L, which is below human detection level defined at 5 ng/L. This geosmin prediction map did not correspond to the *Pseudanabaena spp*. nor *Planktothrix agardii* biovolume distribution map. The selected grid cell with known field collected water quality data located in the ECR north basin used for comparison of observed and predicted values had a higher concentrations of geosmin on September 30, 2008 at 11.35 ng/L than what the model simulation indicated (3.43 ng/l) showing the model's underprediction. The conclusions based on this prediction and the whole reservoir geosmin concentration simulation could be premature, considering the other error factors such as water quality simulation results.

The EFDC water quality model was developed based on estimated SWAT watershed model, since there was no other sufficient available data. The significant boundary condition influence on water quality simulation output was apparent and the differences between actual and simulated SWAT constituents couldn't be effectively minimized during the calibration process. The effects of using estimated watershed inputs were

reflected in high relative error statistics for the EFDC simulation. The EFDC model had also other data insufficiencies, which included solar radiation source location and estimated water temperature for the watershed. Since the water quality simulation output was directly linked to ANFIS models, it had an important impact on overall ANFIS prediction output. Out of two metabolite models ANFIS 2-MIB and ANFIS Geosmin, only one (ANFIS-Geosmin) was confirmed and validated. The 2-MIB model prediction was not validated as it produced out of range values. This could be a result of water quality prediction for ANFIS 2-MIB model variables, but also due to ANFIS preliminary training and testing.

The ANFIS models were trained and tested on a small amount of data of total 112 data pairs. To allow the training, the 102 data pairs had to be used for the amount of variables simulated, which allowed only 10 data pairs for model validation. The ten data pairs is not enough when validating the model and more data is needed in the future for model development. The ANFIS 2-MIB model final validation was poor due to two data points significant over-prediction. This model further was found not applicable, when combined with EFDC prediction variables. This effect is a combined result of multiple errors due to data insufficiency in model development.

Also, the ANFIS models for algal species *Pseudanabaena spp.* and *Planktothrix agardhii* were designed to predict the biovolumes to better communicate the results as most of the available algal research reports the algal counts in cell/mL or uses biovolumes (mm^3/m^3). The algal results from EFDC are given as concentration mg C/L. To be able to use these results, data had to be transformed using the estimate of algal species carbon contribution at the time. This calculation requires a general knowledge of species composition in the reservoir. This not only limits the developed methodology, but also allows for error propagation. In the future it would be recommended to use concentrations of mg C/L for algal data used in ANFIS modeling to avoid the estimations of carbon contributions from algal species, as this knowledge is limited only to selected sampled locations, which may be not representative of the whole reservoir.

Addressing these issues and sources of errors may improve the model predictive capability and allow on its wide application in algal metabolite predictions, considering the heretic often unknown nature of the mechanisms, which dictate the metabolite release from algal cells.

This study was focused on designing a predictive tool, which could be capable of providing the geosmin and 2-MIB concentrations based on available data. The algal species interactions and internal mechanisms leading to release of the metabolite were not explored as they are still not well understood and described in existing literature.

In conclusion of this research, it was confirmed that the metabolite concentration variation within water column has a seasonal character interconnected with the water stability, which was also previously indicated by other researchers (Westerhoff et al., 2002, 2005; Ligor, Buszewski, 2005; Taylor et al., 2006; Journey et al., 2008) and climatic conditions. Both of the taste and odor metabolites were correlated with temperature. However, phosphorus recognized by Dzialowski et al. (2007) as a strong geosmin predictor was not confirmed in this study. Phosphorus and nitrogen were both among predictive variables for *Pseudanabaena spp.* and *Planktothrix agardhii*, which signifies the importance of external and internal loadings into the reservoir and their influence on algal blooms.

The predicted spatial distribution of selected algal species and the metabolite concentrations was not correlated as it was originally assumed, but indicated tributary inflows and shallow parts of the reservoir as areas of increased geosmin concentrations. The analysis of the spatial distribution of algal by-products based on models developed here could be further used for determination of sources of 2-MIB and geosmin. This capability makes current research significant as it could allow on effective treatment of the reservoir and lead to development of better watershed practices if upstream sources of metabolites are detected.

Station ID	Easting	Northing	Sample ID	Sample Number	Sample Date
ECRAT-A3	4412353	559314	ECRAT-A3-C	0805-509	5/22/2008
ECRAT- B3	4411351	559316	ECRAT-B2-C	0806-533	6/5/2008
ECRAT-C4	4408879	559405	ECRAT-C4-C	0806-549	6/17/2008
ECRAT-D1	4408721	559554	ECRAT-D1-C	0807-577	7/8/2008
ECRAT-E1	4408719	559526	ECRAT-E1-C	0807-601	7/16/2008
ECRAT-E3	4412618	559332	ECRAT-E3-C	0807-605	7/16/2008
ECRAT-F1	4408808	559438	ECRAT-F1-C	0807-617	7/30/2008
ECRAT-F3	4412557	559308	ECRAT-F3-C	0807-621	7/30/2008
ECRAT-G1	4408716	559482	ECRAT-G1-C	0808-519	8/14/2008
ECRAT-G3	4412793	559274	ECRAT-G3-C	0808-523	8/14/2008
ECRAT-H1	4408770	559494	ECRAT-H1-C	0808-535	8/20/2008
ECRAT-H3	4412584	559257	ECRAT-H3-C	0808-539	8/20/2008
ECRAT-I1	4408790	559454	ECRAT-I1-C	0808-556	8/27/2008
ECRAT-I3	4472571	559340	ECRAT-I3-C	0808-561	8/27/2008
ECRAT-J4	4408685	559407	ECRAT-J4-C	0809-589	9/3/2008
ECRAT-K1	4408632	559477	ECRAT-K1-C	0809-613	9/16/2008
ECRAT-K3	4412597	559321	ECRAT-K3-C	0809-618	9/16/2008
ECRAT-L1	4408694	559559	ECRAT-L1-C	0809-645	9/30/2008
ECRAT-L3	4412667	559341	ECRAT-L3-C	0809-650	9/30/2008
ECRAT-M1	4408702	559557	ECRAT-M1-C	0810-523	10/16/2008
ECRAT-M3	4412883	559233	ECRAT-M3-C	0810-528	10/16/2008
ECRAT-N1	4418744	559516	ECRAT-N1-C	0810-553	10/28/2008
ECRAT-N3	4412879	559289	ECRAT-N3-C	0810-557	10/28/2008

Table 1: List of water quality monitoring stations including algal information (2008).

Sample Location	Northing	Easting	Sample ID	Date	Sample Location	Northing	Easting	Sample ID	Date
A1	559591	4408835	0904-517	04/22/09	H3	559241	4412782	0907-534	07/21/09
A2	559285	4411120	0904-519	04/22/09	I1	559580	4408700	0908-518	08/06/09
A3	559326	4412780	0904-521	04/22/09	I2	559292	4411268	0908-520	08/06/09
B1	559537	4408702	0905-503	05/04/09	I3	559223	4412781	0908-522	08/06/09
B2	559249	4411115	0905-505	05/04/09	J1	559566	4408679	0908-560	08/18/09
B3	559257	4412786	0905-507	05/04/09	J2	559299	4411165	0908-562	08/18/09
C1	559514	4408757	0905-523	05/19/09	J3	559244	4412779	0908-564	08/18/09
C2	559245	4411187	0905-525	05/19/09	K1	559531	4408672	0909-504	09/02/09
C3	559285	4412867	0905-527	05/19/09	K2	559262	4411300	0909-506	09/02/09
D1	559542	4408655	0906-503	06/02/09	K3	559247	4412776	0909-508	09/02/09
D2	559274	4411138	0906-505	06/02/09	L1	586896	4418358	0909-578	09/17/09
D3	559229	4412735	0906-507	06/02/09	L2	559231	4411103	0909-580	09/17/09
E1	559610	4408725	0906-524	06/17/09	L3	559229	4412768	0909-582	09/17/09
E2	559274	4411109	0906-526	06/17/09	M1	559547	4408690	0910-500	10/01/09
E3	559256	4412789	0906-528	06/17/09	M2	559280	4411198	0910-502	10/01/09
F1	559556	4408704	0907-500	07/01/09	M3	559245	4412782	0910-504	10/01/09
F2	559265	4411101	0907-502	07/01/09	N1	559552	4408668	0910-545	10/15/09
F3	559243	4412776	0907-504	07/01/09	N2	559234	4411129	0910-547	10/15/09
G1	559533	4408712	0907-514	07/09/09	N3	559223	4412787	0910-549	10/15/09
G2	559220	4411100	0907-516	07/09/09	O1	559565	4408723	0910-565	10/28/09
G3	559227	4412775	0907-518	07/09/09	O2	559237	4411133	0910-567	10/28/09
H1	559576	4408686	0907-530	07/21/09	O3	559261	4412784	0910-569	10/28/09
H2	559229	4411137	0907-532	07/21/09					

Table 2: List of water quality monitoring stations including algal information (2009).

Station ID	Easting	Northing	Station ID	Sample Date	Station ID	Easting	Northing	Station ID	Sample Date
A1	4408685	559598	1004-524	4/21/2010	I1	4408698	559553	1007-530	7/13/2010
A2	4411135	559348	1004-526	4/21/10	I2	4411122	559247	1007-532	7/13/10
A3	4412750	559267	1004-528	4/21/2010	I3	4412789	559261	1007-537	7/13/2010
B1	4408679	559566	1005-500	5/4/0210	J1	4408701	559565	1007-591	7/28/2010
B2	4411131	559335	1005-502	5/4/10	J2	4411120	559257	1007-593	7/28/10
B3	4412644	559330	1005-507	5/4/0210	J3	4412788	559261	1007-598	7/28/2010
C1	4408694	559557	1005-553	5/18/2010	K1	4408673	559586	1008-523	8/10/2010
C2	4411058	559306	1005-555	5/18/10	K2	4411117	559288	1008-525	8/10/10
C3	4412721	559262	1005-560	5/18/2010	K3	4412745	559261	1008-530	8/10/2010
D1	4408692	559540	1005-578	5/24/2010	L1	4408662	559541	1008-579	8/24/2010
D3	4412813	559254	1005-585	5/24/2010	L2	4411106	559250	1008-581	8/24/10
E1	4408713	559581	1005-587	5/27/2010	L3	4412789	559254	1008-586	8/24/2010
E3	4412680	559256	1005-591	5/27/2010	M1	4408716	559606	1009-500	9/8/2010
F1	4408723	559608	1006-517	6/3/2010	M2	4411114	559261	1009-502	9/8/10
F2	4411194	559295	1006-519	6/3/10	M3	4412758	559251	1009-507	9/8/2010
F3	4412823	559290	1006-524	6/3/2010	N1	4408744	559549	1009-554	9/21/2010
G1	4408749	559595	1006-548	6/15/2010	N2	4411116	559246	1009-556	9/21/10
G2	4411167	559240	1006-550	6/15/10	N3	4412789	559248	1009-561	9/21/2010
G3	4412871	559287	1006-555	6/15/2010	O1	4408702	559557	1009-900	9/30/2010
H1	4408701	559519	1007-500	7/1/2010	O3	4412758	559256	1009-904	9/30/2010
H2	4411128	559242	1007-502	7/1/10	P1	4408686	559575	1010-500	10/4/2010
H3	4412790	559242	1007-507	7/1/2010	P3	4412758	559256	1010-507	10/4/2010

Table 3: List of water quality monitoring stations including algal information (2010).

Variable	N of Cases	Min	Max	Mean	STDE	Shapiro-Wilk Statistics	Shapiro-Wilk p-Value
CYANOPHYTES	112	0.72	6.97	3.886	1.469	0.921	0.000
DIATOMS	112	0	6.58	3.55	1.47	0.929	0.000
PSEUDANABAENA spp.	112	0	6.72	3.391	1.695	0.939	0.000
PLANKTROTHRIX AGARDHII	112	0	5.67	1.657	1.486	0.843	0.000
CYLINDROSPERMOPSIS RACIBORSKI	112	0	5.25	1.237	1.55	0.76	0.000
CHLOROPHYLL -a	112	0	3.06	1.521	0.432	0.953	0.001
TOTAL DEPTH	112	0.61	1.12	0.885	0.16	0.888	0.000
PHOTIC DEPTH	112	0.01	0.8	0.479	0.114	0.965	0.005
WATER TEMPERATURE	112	1.05	1.48	1.353	0.1	0.888	0.000
SPC	112	0.13	0.21	0.168	0.02	0.965	0.005
CONDUCTIVITY	112	0.13	0.26	0.16	0.019	0.884	0.000
SALINITY	112	0.07	0.12	0.089	0.011	0.915	0.000
DO	112	0.8	1.42	1.063	0.106	0.991	0.687
PH	112	0.89	1	0.97	0.014	0.861	0.000
ORP	112	1.61	2.44	2.172	0.164	0.899	0.000
MICROCYSTIN	112	0.01	0.42	0.065	0.058	0.669	0.000
SILICA	112	0.12	1.05	0.565	0.187	0.987	0.386
ALKALINITY	112	2.04	2.25	2.144	0.054	0.965	0.005
CL	112	1.34	1.85	1.626	0.108	0.977	0.049
SO4	112	0.48	1.56	1.391	0.147	0.62	0.000
TOTAL P	112	0.05	0.29	0.034	0.03	0.555	0.000
NO2	112	0.01	0.1	0.024	0.016	0.794	0.000
NO3	112	0.04	0.66	0.26	0.197	0.886	0.000
TKN	112	0.04	0.49	0.333	0.072	0.895	0.000
TOTAL N	112	0.11	0.78	0.487	0.143	0.987	0.365
NH3-N	112	0.02	0.22	0.03	0.03	0.703	0.000
Ca	112	1.23	1.83	1.647	0.094	0.927	0.000
Mg	112	0.88	1.41	1.252	0.07	0.916	0.000
K	112	0.42	0.69	0.603	0.048	0.957	0.001
Na	112	1.01	1.61	1.407	0.117	0.971	0.016
TOTAL HARDNESS	112	1.83	2.38	2.254	0.077	0.883	0.000

Variable	N of Cases	Min	Max	Mean	STDE	Shapiro-Wilk Statistics	Shapiro-Wilk p-Value
FLUORIDE	112	0.01	0.2	0.072	0.02	0.828	0.000
RTRM MAX. VALUE	112	0	2.03	1.296	0.499	0.932	0.000
MAX. VALUE RTRM DEPTH	112	0.18	1.11	0.692	0.236	0.949	0.000
SUM Ri	112	-1.05	3.06	1.442	0.862	0.979	0.078
AVERAGE DIFFUSION COEFFICIENT	112	-6.62	-2.03	-4.168	0.997	0.983	0.160
SUM RTRM	112	0	2.49	1.744	0.534	0.944	0.000
PRECIPITATION (MOVING AVERAGE)	112	0	1.21	0.569	0.288	0.969	0.010
GEOSMIN	112	0.35	2.04	0.942	0.431	0.887	0.000
2- MIB	112	0.35	2.35	1.034	0.504	0.929	0.000

Table 4: Shapiro-Wilk normality test results for selected environmental parameters for ECR.

Variable	*Pseudanabaena spp.*	*Planktothrix agardhii*	Geosmin	MIB	Variable	*Pseudanabaena spp.*	*Planktothrix agardhii*	Geosmin	MIB
Cyanophytes	.942**	0.077	0.007	-0.022	**Total P**	.418**	0.026	0.121	.319**
	0	0.426	0.94	0.814		0	0.788	0.202	0.001
Diatoms	.719**	-.225*	-.194*	-0.06	**NO2**	-0.109	-0.134	-0.164	.301**
	0	0.017	0.04	0.532		0.254	0.159	0.085	0.001
Pseudanabaena spp.	1	0.003	-0.018	0.103	**NO3**	-.423**	-.232*	-0.122	-0.075
	.	0.975	0.848	0.278		0	0.014	0.199	0.43
Planktothrix agardhii	0.003	1	.622**	.239**	**TKN**	-0.031	.195*	.256**	0.057
	0.975	.	0	0.011		0.743	0.039	0.007	0.553
Cylindrospermopsis raciborskii	.219*	.272**	.230*	0.013	**Total N**	-.404**	-0.179	-0.019	-0.094
	0.02	0.004	0.015	0.894		0	0.059	0.839	0.322
Chlorophyll -a	-.194*	.315**	0.121	-0.033	**NH3-N**	-.213*	0.032	-0.001	.219*
	0.04	0.001	0.203	0.733		0.024	0.737	0.993	0.02
Total Depth	0.022	0.046	-0.016	-0.006	**Ca**	-.592**	-0.091	0.039	0.136
	0.815	0.632	0.867	0.949		0	0.339	0.686	0.152
Photic Depth	-.291**	-0.083	-0.142	-.214*	**Mg**	-.212*	.363**	.308**	.284**
	0.002	0.385	0.138	0.023		0.025	0	0.001	0.002
Water Temperature	.232*	-0.096	-.345**	-.411**	**K**	.469**	-.233*	-.308**	-0.027
	0.014	0.312	0	0		0	0.013	0.001	0.776
SPC	-.343**	.280**	.307**	.351**	**Na**	0.155	.347**	.273**	.551**
	0	0.003	0.001	0		0.102	0	0.004	0
Conductivity	-.234*	.214*	0.111	0.154	**Total Hardness**	-.526**	0.033	0.129	.231*
	0.013	0.024	0.244	0.104		0	0.727	0.174	0.014
TDS	-.341**	.292**	.310**	.355**	**Fluoride**	.243**	0.094	0.033	.241*
	0	0.002	0.001	0		0.01	0.325	0.729	0.01
Salinity	-.329**	.289**	.300**	.354**	**RTRM (max. value)**	0.045	0.017	-0.105	-0.172
	0	0.002	0.001	0		0.64	0.86	0.271	0.07
DO	0.174	0.058	0.136	-0.052	**Max. Value RTRM Depth**	-0.099	-0.132	-0.121	-0.053
	0.066	0.542	0.152	0.585		0.301	0.165	0.204	0.578
Ph	.311**	0.049	-0.026	-0.183	**Sum Ri**	0.011	0.006	-0.108	-.258**
	0.001	0.61	0.788	0.053		0.908	0.953	0.259	0.006
ORP	0.174	0.107	0.151	0.004	**Average Diffusion**	-0.052	-0.039	0.108	.271**
	0.066	0.263	0.111	0.963		0.589	0.685	0.258	0.004
Microcystin	0.051	.588**	.590**	.480**	**Sum RTRM**	0.025	-0.034	-0.13	-.238*
	0.595	0	0	0		0.792	0.725	0.171	0.012
Silica	0.126	-0.103	-0.099	0.075	**Precipitation**	-.283**	0.087	0.121	-0.131

Variable	***Pseudanabaena spp.***	***Planktothrix agardhii***	Geosmin	MIB	Variable	***Pseudanabaena spp.***	***Planktothrix agardhii***	Geosmin	MIB
	0.187	0.281	0.3	0.432		0.003	0.361	0.205	0.167
Alkalinity	-.549**	-0.023	0.115	.297**	**Geosmin**	-0.018	.622**	1	.299**
	0	0.814	0.227	0.001		0.848	0	.	0.001
Cl-	0.079	.397**	.338**	.577**	**MIB**	0.103	.239**	.299**	1
	0.406	0	0	0		0.278	0.011	0.001	.
SO4	-.212*	.385**	.324**	.360**					
	0.025	0	0	0					

Table 5: Spearman's rho correlation results

(first raw for variable gives correlation coefficient and second 2-tailed significance).

Focal Point	Number of grids	Minimum cell size (m)	Maximum cell size (m)	Expanding factor	Cell rotation
(X, Y)					angle (degree)
559313; 4411545	5290	15.5	100	1.02	-52
559313; 4411545	3821	20	100	1.02	-52
559313; 4411545	1763	35	80	1.02	-52
559313; 4411545	**2401**	**40**	**60**	**1.005**	**0**

Table 6: Grid generation variants. Final grid for ECR is marked in bold.

Boundary Group	**EFDC Assigned Cells (I,J)**
Eagle Creek tributary (110)	(35,134),(36,134),(37,135),(38,135),(39,135)
111	(34,125)
Fishback Creek (112)	(32,121)
113	(29,110)
Bush Creek (114)	(51,111)
115	(27,92)
116	(54,91),(54,90)
117	(27,84),(26,83)
118	(21,64)
119	(22,50)
120	(28,38)
School Branch (127)	(3,29),(4,30)
Dam	(43,5),(44,5)
Intake	(44,43)

Table 7: ECR grid cell assignment for tributary inflows.

Temperature (^{0}C)	Average	Average Min.	Average Max.	Daily Average					
				Low			High		
				Ave	Min	Max	Ave	Low	Max
2008	11	-13	28	6.2	-17	23	16.3	-11	33
2009	12	-18	28	6.6	-25	23	16.2	-13	33
2010	12	-13	31	7.1	-17	25	17	-8	35

Precipitation (mm)				Spring			Summer			Fall		
				Ave	Min	Max	Ave	Min	Max	Ave	Min	Max
2008	3.3	0	77.47	5.49	0	77.4	2.79	0	41.4	1.78	0	34
2009	3	0	95.75	5.47	0	63.5	2.64	0	95.7	2.72	0	32.2
2010	2.28	0	48.76	4.5	0	48.7	2.42	0	37.8	1.80	0	35.8

Table 8: Selected climatic data summary for ECR (2008-2010).

Marina Stations	Date	Net Growth Rate (d^{-1})	Dam Stations	Date	Net Growth Rate (d^{-1})	Intake Stations	Date	Net Growth Rate (d^{-1})
A3	5/22/08	0.39	O3	9/30/10	0.40	M1	9/8/10	0.52
B3	6/5/08	0.44	P3	10/4/10	0.29	N1	9/21/10	0.49
E3	7/16/08	0.68	C4	6/17/08	0.56	O1	9/30/10	0.42
F3	7/30/08	0.67	D1	7/8/08	0.52	P1	10/4/10	0.34
G3	8/16/08	0.58	E1	7/16/08	0.68	A2	4/22/09	0.18
H3	8/20/08	0.61	F1	7/30/08	0.65	B2	5/4/09	0.29
I3	8/27/08	0.57	G1	8/16/08	0.59	C2	5/19/09	0.34
K3	9/16/08	0.46	H1	8/20/08	0.62	D2	6/2/09	0.48
L3	9/30/08	0.43	I1	8/27/08	0.59	E2	6/17/09	0.49
M3	10/14/08	0.39	K1	9/16/08	0.47	F2	7/1/09	0.60
N3	10/28/08	0.17	L1	9/30/08	0.44	G2	7/9/09	0.53
A3	4/22/09	0.20	M1	10/14/08	0.37	H2	7/21/09	0.55
B3	5/4/09	0.31	N1	10/28/08	0.21	I2	8/6/09	0.55
C3	5/19/09	0.34	A1	4/22/09	0.18	J2	8/18/09	0.67
D3	6/2/09	0.49	B1	5/4/09	0.33	K2	9/2/09	0.51
E3	6/17/09	0.51	C1	5/19/09	0.32	L2	9/17/09	0.51
F3	7/1/09	0.59	D1	6/2/09	0.46	M2	10/1/09	0.38
G3	7/9/09	0.53	E1	6/17/09	0.44	N2	10/15/09	0.24
H3	7/21/09	0.58	F1	7/1/09	0.59	O2	10/28/09	0.21
I3	8/6/09	0.57	G1	7/9/09	0.55	A2	4/21/10	0.28
J3	8/18/09	0.69	H1	7/21/09	0.55	B2	5/4/10	0.33
K3	9/2/09	0.52	I1	8/6/09	0.57	C2	5/18/10	0.32
L3	9/17/09	0.51	J1	8/18/09	0.65	F2	6/3/10	0.57
M3	10/1/09	0.34	K1	9/2/09	0.51	G2	6/15/10	0.61
N3	10/15/09	0.22	L1	9/17/09	0.53	H2	7/1/10	0.60
O3	10/28/09	0.21	M1	10/1/09	0.39	I2	7/13/10	0.73
A3	4/21/10	0.29	N1	10/15/09	0.25	J2	7/28/10	0.80
B3	5/4/10	0.35	O1	10/28/09	0.21	K2	8/10/10	0.75
C3	5/18/10	0.33	A1	4/21/10	0.31	L2	8/24/10	0.68
D3	5/24/10	0.41	B1	5/4/10	0.32	M2	9/8/10	0.53
E3	5/27/10	0.48	C1	5/18/10	0.33	N2	9/21/10	0.51
F3	6/3/10	0.62	D1	5/24/10	0.39			
G3	6/15/10	0.59	E1	5/27/10	0.46			
H3	7/1/10	0.62	F1	6/3/10	0.60			

Marina Stations	Date	Net Growth Rate (d^{-1})	Dam Stations	Date	Net Growth Rate (d^{-1})	Intake Stations	Date	Net Growth Rate (d^{-1})
I3	7/13/10	0.76	G1	6/15/10	0.60			
J3	7/28/10	0.80	H1	7/1/10	0.64			
K3	8/10/10	0.77	I1	7/13/10	0.75			
L3	8/24/10	0.70	J1	7/28/10	0.78			
M3	9/8/10	0.51	K1	8/10/10	0.78			
N3	9/21/10	0.53	L1	8/24/10	0.70			

Table 9: Cyanobacteria estimated growth rate calculations for monitoring stations in three regions (marina, dam, and intake) according to Wong et al. (2009).

Pseudanabaena spp. **(mm^3/m^3)**	**Marina**	**Dam**	**Intake**	***Planktrothrix agardhii*** **(mm^3/m^3)**	**Marina**	**Dam**	**Intake**
2008				**2008**			
MIN.	82.94	212.06	N/A	MIN.	0	0	N/A
MAX.	5,202.48	5,089.38	N/A	MAX.	471.24	70.69	N/A
PEAK DATE	8/20/08	8/20/08	N/A	PEAK DATE	7/30/08	8/16/08	N/A
2009				**2009**			
MIN.	0	0	0	MIN.	0	0	0
MAX.	4,665.27	2,172.41	3,689.80	MAX.	4,594.58	3,828.82	4,123.34
PEAK DATE	10/1/09	9/17/09	10/1/09	PEAK DATE	10/15/09	10/28/09	10/28/09
2010				**2010**			
MIN.	494.8	268.61	254.47	MIN.	0	0	0
MAX.	12,668.47	10,532.19	7,791.15	MAX.	1,060.29	2,427.00	848.00
PEAK DATE	8/10/2010	6/3/2010	8/10/2010	PEAK DATE	5/18/2010	5/24/2010	4/21/2010

Table 10: *Pseudanabaena spp.* and *Planktrothrix agardhii* biovolumes

for 3 season period in 2008-2010.

	2008				2009				2010			
	Min.	Max.	Mean	STD	Min.	Max.	Mean	STD	Min.	Max.	Mean	STD
Geosmin	1.25	11.35	4.46	2.78	1.25	80.43	16.46	21.96	1.25	**109.43**	18.88	26.3
2- MIB	2.5	29.22	6.95	6.71	1.25	18.43	7.98	5.13	1.25	**223.72**	46.91	56.9

Table 11: Geosmin and 2-MIB statistics for 2008-2010 ECR.

Station ID	Profile Date	# Pairs	Relative Error	RMSE Error	Data Average	Model Average
ECRAT-A1	5/22/08	9	2.99	0.57	16.32	16.04
ECRAT-A2	5/22/08	8	5.44	0.96	15.96	16.38
ECRAT-A3	5/22/08	6	6.96	1.18	16.27	17.40
ECRAT-A4	5/22/08	5	21.93	3.93	15.95	19.45
ECRAT-B1	6/5/08	11	10.48	2.18	19.20	20.97
ECRAT-B2	6/5/08	10	3.29	0.88	20.82	20.42
ECRAT-B3	6/5/08	7	3.30	0.70	20.70	21.39
ECRAT-B4	6/5/08	5	1.44	0.30	21.01	21.32
ECRAT-C1	6/17/08	5	7.19	1.81	22.35	23.96
ECRAT-C2	6/17/08	5	10.01	2.37	23.49	25.85
ECRAT-C3	6/17/08	8	9.94	2.44	23.09	25.39
ECRAT-C4	6/17/08	13	2.92	0.92	22.12	22.69
ECRAT-D1	7/8/08	13	7.71	2.10	22.22	23.62
ECRAT-D2	7/8/08	9	2.01	0.56	24.17	23.99
ECRAT-D3	7/8/08	5	4.24	1.10	24.79	25.29
ECRAT-D4	7/8/08	4	1.52	0.40	24.98	24.81
ECRAT-E1	7/16/08	12	3.05	0.89	23.56	23.68
ECRAT-E2	7/16/08	7	4.23	1.25	25.90	26.19
ECRAT-E3	7/16/08	7	3.22	1.11	26.52	27.36
ECRAT-E4	7/16/08	3	3.84	1.14	27.84	28.56
ECRAT-F1	7/30/08	14	3.60	0.99	23.77	23.58
ECRAT-F2	7/30/08	9	3.69	1.16	26.05	25.51
ECRAT-F3	7/30/08	6	3.37	1.02	26.85	27.14
Average Statistics		**181**	**5.27**	**1.28**		

Table 12: Temperature calibration error statistics.

Station ID	Profile Date	# Pairs	Relative Error	RMS Error	Data Average	Model Average
ECRAT-G1	8/14/08	11	2.94	0.93	23.83	23.48
ECRAT-G2	8/14/08	8	4.22	1.16	24.85	24.58
ECRAT-G3	8/14/08	6	4.53	1.15	24.88	26.01
ECRAT-H1	8/20/08	12	9.47	2.44	24.07	21.79
ECRAT-H2	8/20/08	7	8.64	2.33	25.02	22.86
ECRAT-H3	8/20/08	6	9.30	2.53	25.43	23.06
ECRAT-H4	8/20/08	4	6.76	2.06	25.45	23.98
ECRAT-I1	8/27/08	14	7.49	2.04	23.36	21.67
ECRAT-I2	8/27/08	8	6.10	1.86	24.40	23.76
ECRAT-I3	8/27/08	7	6.06	1.58	24.28	25.58
ECRAT-I4	8/27/08	5	18.31	4.38	23.57	27.88
ECRAT-J1	9/3/08	7	6.79	2.25	26.28	27.60
ECRAT-J2	9/3/08	8	5.13	1.57	26.15	26.57
ECRAT-J3	9/3/08	12	6.65	1.99	25.08	24.79
ECRAT-J4	9/3/08	16	9.24	2.60	25.24	22.91
ECRAT-K1	9/16/08	13	6.26	1.64	21.86	20.54
ECRAT-K2	9/16/08	7	2.03	0.49	22.27	21.82
ECRAT-K3	9/16/08	6	0.54	0.16	21.79	21.68
ECRAT-K4	9/16/08	4	4.92	1.11	21.19	22.23
ECRAT-L1	9/30/08	11	6.45	1.79	21.25	19.88
ECRAT-L2	9/30/08	8	2.37	0.66	21.45	21.23
ECRAT-L3	9/30/08	6	5.19	1.13	21.06	22.15

Station ID	Profile Date	# Pairs	Relative Error	RMS Error	Data Average	Model Average
ECRAT-L4	9/30/08	4	12.92	2.67	20.31	22.94
ECRAT-M1	10/16/08	10	4.44	0.89	18.63	17.80
ECRAT-M2	10/16/08	6	2.48	0.53	19.08	19.55
ECRAT-M3	10/16/08	4	3.37	0.74	19.57	19.60
ECRAT-M4	10/16/08	3	1.53	0.36	20.24	19.93
ECRAT-N1	10/28/08	10	11.51	1.47	12.72	14.18
ECRAT-N2	10/28/08	6	6.80	0.85	12.54	13.39
ECRAT-N3	10/28/08	4	8.94	0.92	10.22	11.13
ECRAT-N4	10/28/08	2	19.73	1.80	9.10	10.89
Composite		**235**	**6.61**	**1.62**		

Table 13: Temperature validation error statistics.

Station	Parameter	Type Date	RMS Error	Data Average	Model Average
Marina	Cyanobacteria (mg/l as C)	Minimum 16-Jul-2008 - 30-Sep-2008	1.307	1.291	0.343
Marina	Cyanobacteria (mg/l as C)	Average 16-Jul-2008 - 30-Sep-2008	1.279	1.291	0.396
Marina	Cyanobacteria (mg/l as C)	Maximum 16-Jul-2008 - 30-Sep-2008	1.257	1.291	0.426
Marina	Diatoms (mg/l as C)	Minimum 05-Jun-2008- 16-Sep-2008	0.23	0.221	0.046
Marina	Diatoms (mg/l as C)	Average 05-Jun-2008 - 16-Sep-2008	0.219	0.221	0.094
Marina	Diatoms (mg/l as C)	Maximum 05-Jun-2008 - 16-Sep-2008	0.324	0.221	0.186
Marina	Green Algae (mg/l as C)	Minimum 05-Jun-2008 - 16-Sep-2008	0.5	0.432	0
Marina	Green Algae (mg/l as C)	Average 05-Jun-2008 - 16-Sep-2008	0.5	0.432	0
Marina	Green Algae (mg/l as C)	Maximum 05-Jun-2008 - 16-Sep-2008	0.5	0.432	0

Composite Statistics		
Parameter		RMS Error Ave. Obs. Ave. Model
Cyanobacteria	mg C/L	1.281 1.291 0.388
Diatoms	mg C/L	0.258 0.221 0.108
Green Algae	mg C/L	0.500 0.432 0.000

Table 14: Algal calibration results for Marina monitoring station.

Station	Parameter	Type Date	RMS Error	Data Average	Model Average
Dam	Cyanobacteria (mg/l as C)	Minimum 08-Jul-2008 - 16-Sep-2008	0.554	0.605	0.176
Dam	Cyanobacteria (mg/l as C)	Average 08-Jul-2008 - 16-Sep-2008	0.519	0.605	0.216
Dam	Cyanobacteria (mg/l as C)	Maximum 08-Jul-2008 - 16-Sep-2008	0.47	0.605	0.27
Dam	Diatoms (mg/l as C)	Minimum 08-Jul-2008 - 16-Sep-2008	0.253	0.229	0.14
Dam	Diatoms (mg/l as C)	Average 08-Jul-2008 - 16-Sep-2008	0.302	0.229	0.178
Dam	Diatoms (mg/l as C)	Maximum 08-Jul-2008 -16-Sep-2008	0.35	0.229	0.209
Dam	Green Algae (mg/l as C)	Minimum 08-Jul-2008- 16-Sep-2008	0.472	0.399	0
Dam	Green Algae (mg/l as C)	Average 08-Jul-2008 - 16-Sep-2008	0.472	0.399	0
Dam	Green Algae (mg/l as C)	Maximum 08-Jul-2008 - 16-Sep-2008	0.472	0.399	0
Composite Statistics					
Parameter	unit	RMS Error	Average Observed	Average Model	
Cyanobact	mg C/L	0.514	0.605	0.221	
Diatoms	mg C/L	0.302	0.229	0.176	

Station	Parameter	Type Date	RMS Error	Data Average	Model Average
Green Algae	mg C/L	0.472	0.399	0	

Table 15: Algal calibration results for Dam monitoring station.

Date	*Pseudanabaena spp.* (mm^3/m^3) Observed	*Pseudanabaena spp.* ($mm^3 \backslash m^3$) Predicted
7/8/2008	814	1348
8/16/2008	2616	1537
9/30/2008	943	1393
5/4/2009	61.32	74
7/1/2009	57.55	60
8/6/2009	217.77	3
9/17/2009	3,408.06	3152
4/21/2010	1,778.00	1625
6/15/2010	1,079.00	6
8/24/2010	3,351.00	4503

Table 16: ANFIS validation results for *Pseudaanabaena spp*..

Date	*Planktrothrix agardhii* Observed (mm^3 /m^3)	*Planktrothrix agardhii* Predicted (mm^3 /m^3)
7/30/2008	1.00	193
8/27/2008	1.00	2
4/22/2009	1.00	53
6/17/2009	1.00	2
7/21/2009	118.81	66
9/17/2009	1,650.34	635
10/28/2009	4,124.34	5,355
6/3/2010	119.00	46
8/24/2010	1.00	32
4/21/2010	566.49	14
9/21/2010	1.00	12

Table 17: ANFIS validation results for *Planktothrix agardhii*.

Date	Geosmin (ng/l) Observed	Geosmin (ng/l) Predicted
7/8/2008	3.9	5.70
8/27/2008	3.62	4.83
5/4/2009	7.79	8.26
7/9/2009	3.88	3.57
9/17/2009	29.22	13.07
5/18/2010	45.35	34.89
8/24/2010	3.26	3.54
7/1/2010	9.41	5.50
9/8/2010	2.25	5.48
10/4/2010	4.14	4.63

Table 18: ANFIS validation results for geosmin.

Date	2-MIB (ng/l) Observed	2-MIB (ng/l) Predicted
7/8/2008	4.87	10.50
8/27/2008	4.69	1.00
5/4/2009	14.09	26.60
7/9/2009	3.76	1.30
9/17/2009	9.27	16.20
5/18/2010	160.81	4181.20
8/24/2010	10.37	4.90
7/1/2010	2.25	5.50
9/8/10	16.04	22.70
10/4/10	25.02	4199.0

Table 19: ANFIS validation results for 2-MIB.

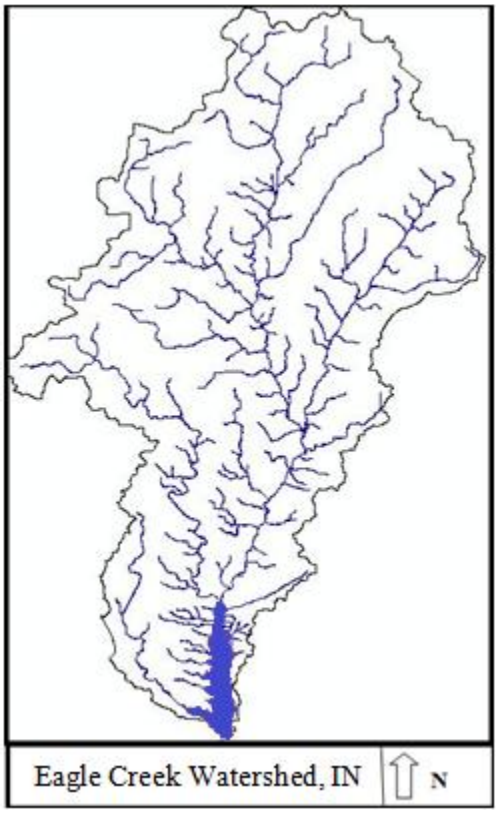


Figure 1: Eagle Creek Watershed, Indiana

(Eagle Creek Reservoir located in the south part of the watershed –solid fill).

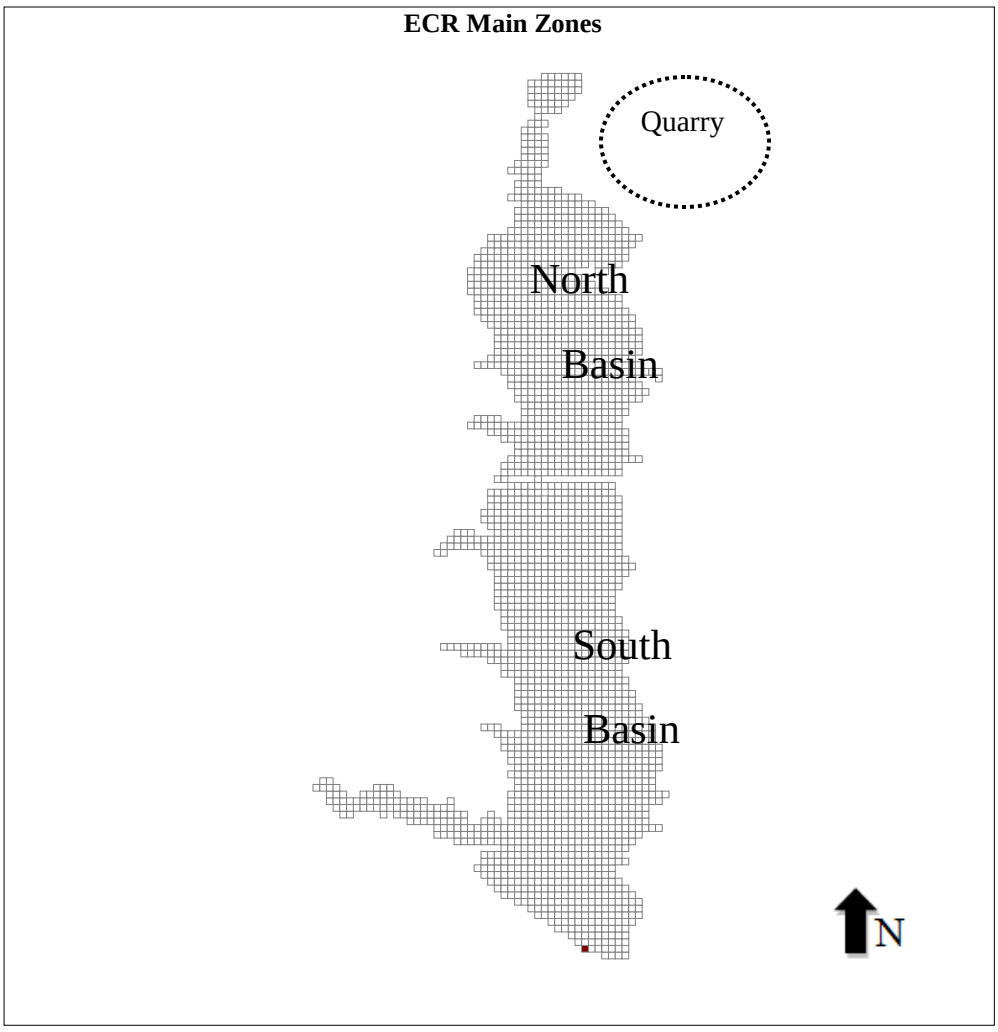


Figure 2: Eagle Creek Reservoir's Basins. Quarry represented by oval dotted shape (not an actual boundary outline) was not included in outgoing research.

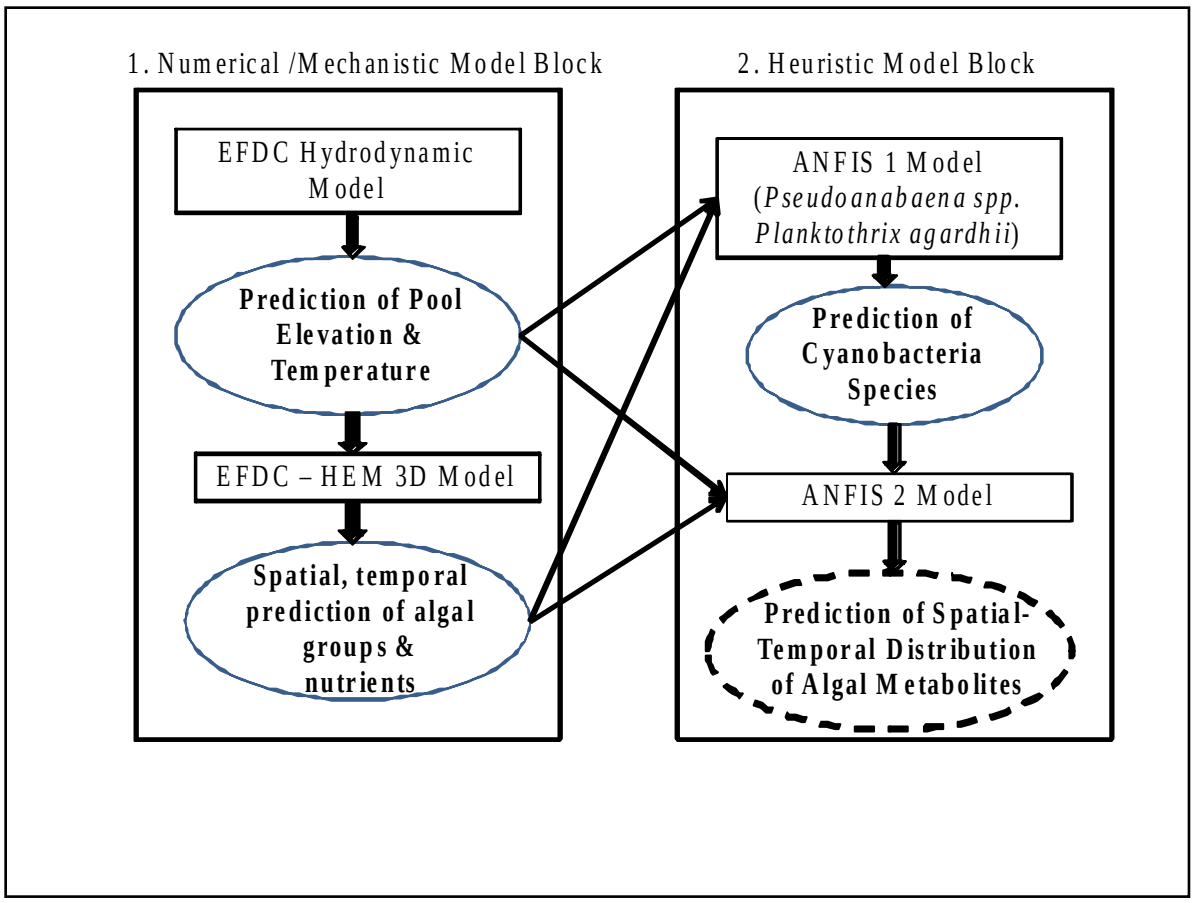


Figure 3: Conceptual model developed for water quality and ANFIS modeling for ECR.

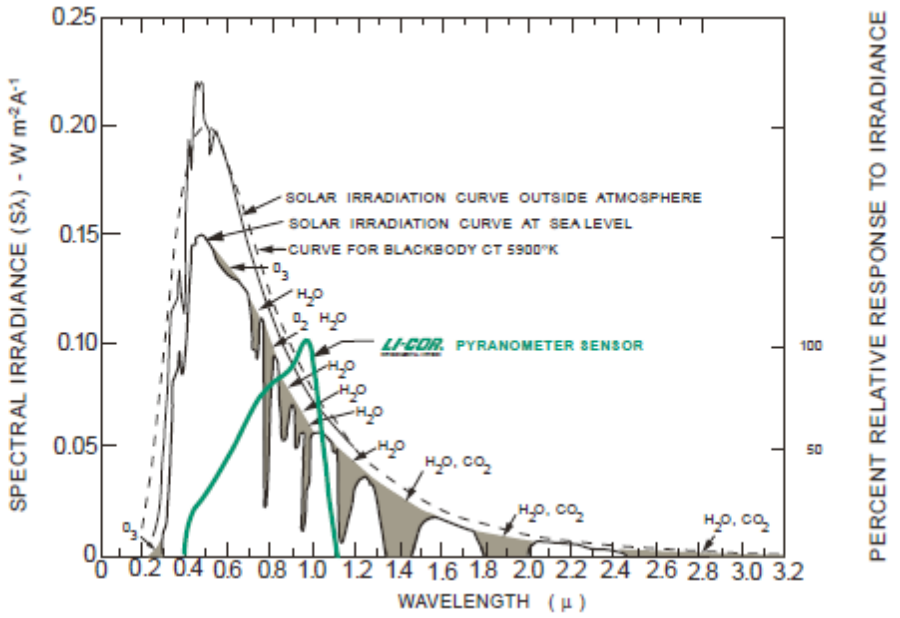


Figure 4: LI-COR pyranometer spectral response and energy distribution in the solar spectrum (Kerr et al.,1967).

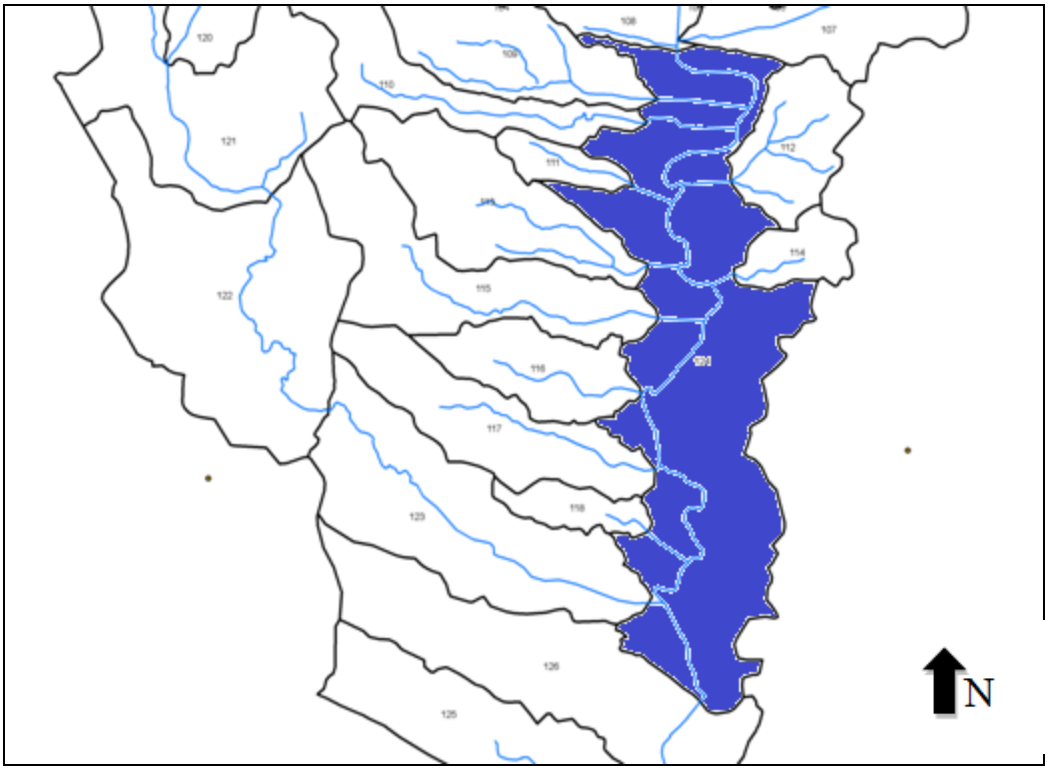


Figure 5: Eagle Creek watershed enumerated sub-basins (selected view).

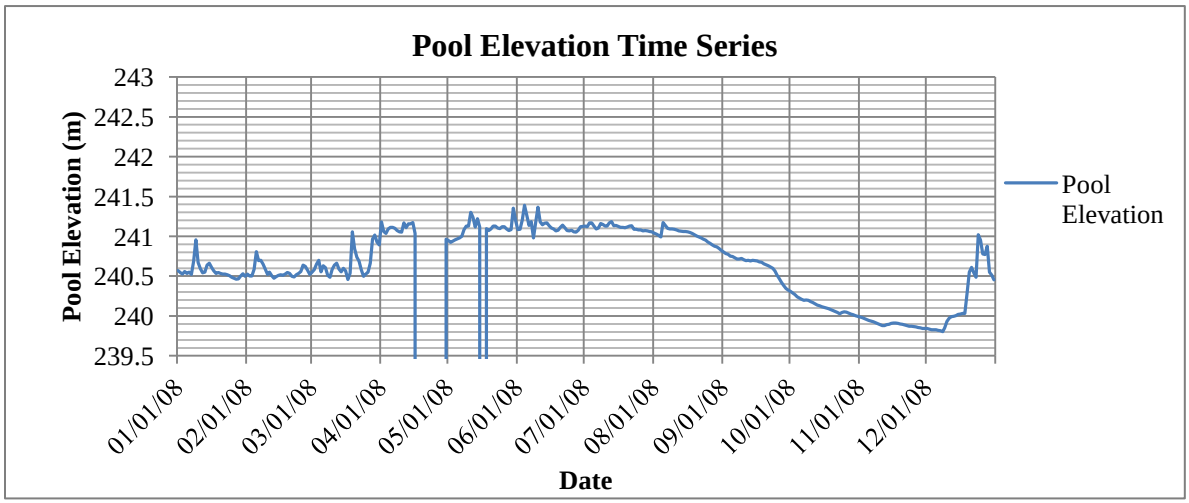


Figure 6: USGS pool elevation measurements for ECR in 2008.

Figure Note:

Sudden drops in observed pool elevation time series were a result of missing data in USGS files for days April 17 -29 and May 16-17. This issue was solved by accepting the value for pool elevation from the last recorded data before the "no data period" until the

day before the next data was available. This decision was justified based on difference calculation between last observed values before no data period and first available recording. The value of 0.07 m difference was considered insignificant.

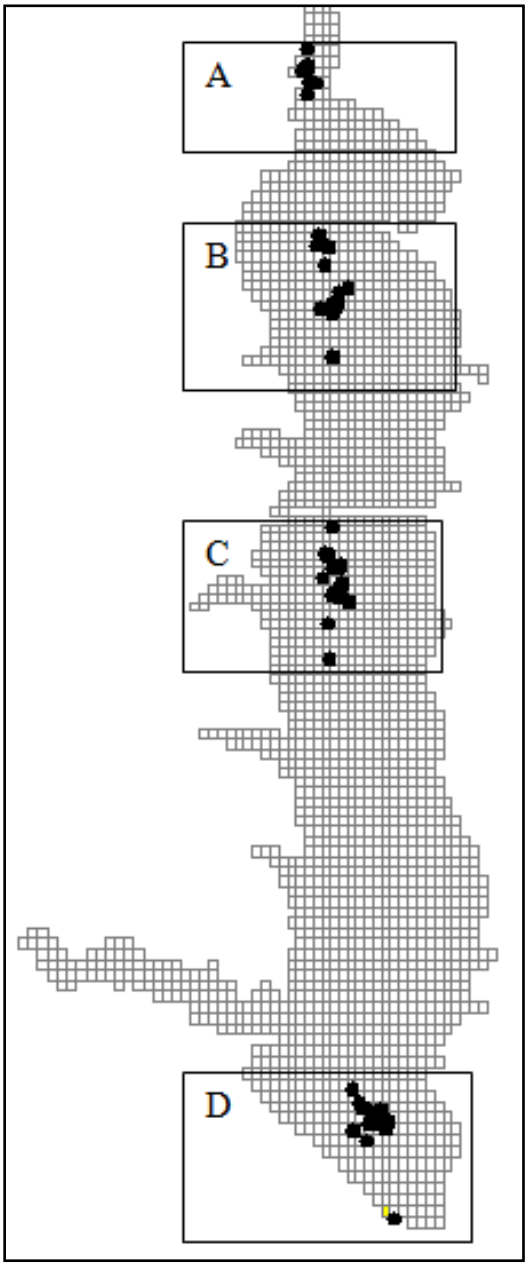


Figure 7: Monitoring stations locations:
A- Eagle Creek region, B- Marina region, C- Dam region, D- Dam region.

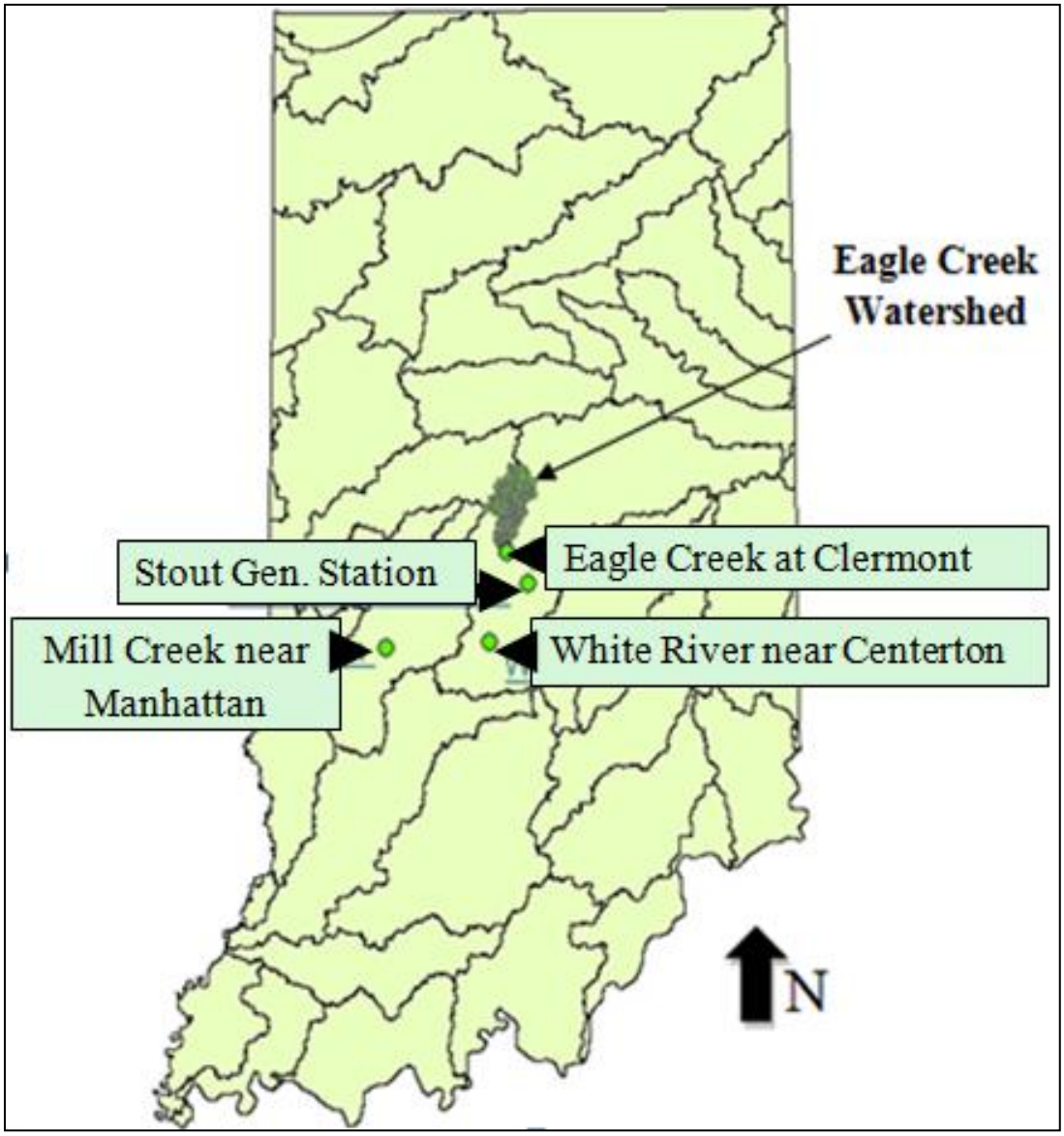


Figure 8: USGS gauge stations near Eagle Creek Reservoir.

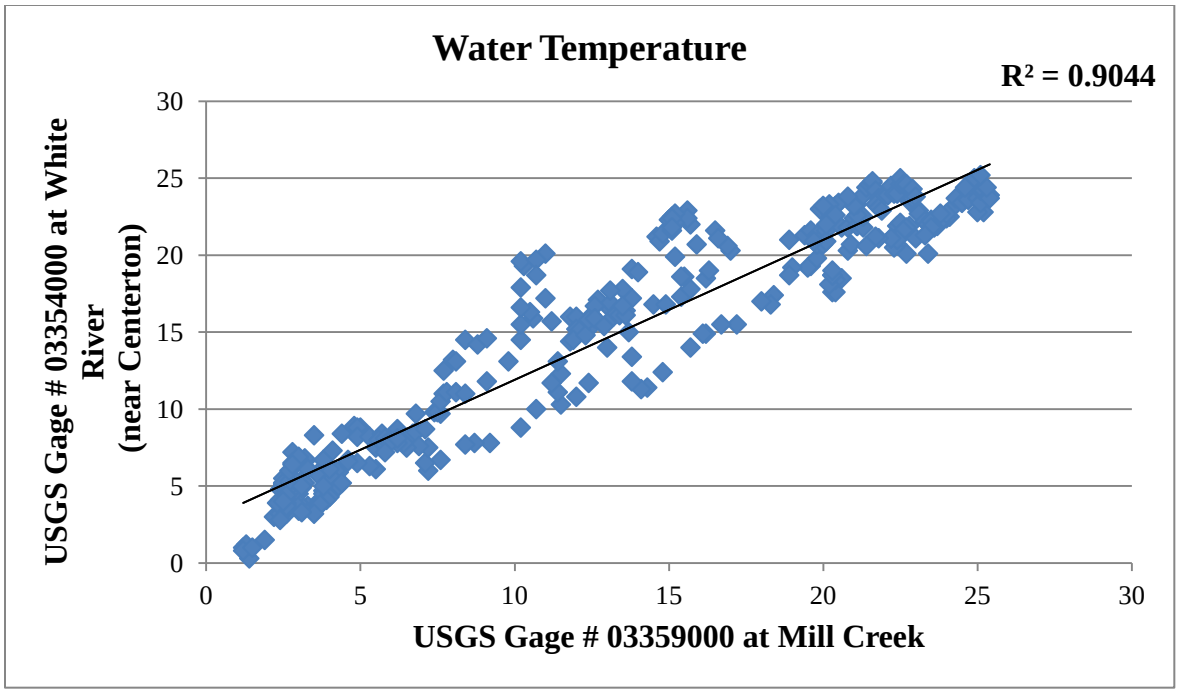


Figure 9: Water temperature regression for selected USGS gage stations.

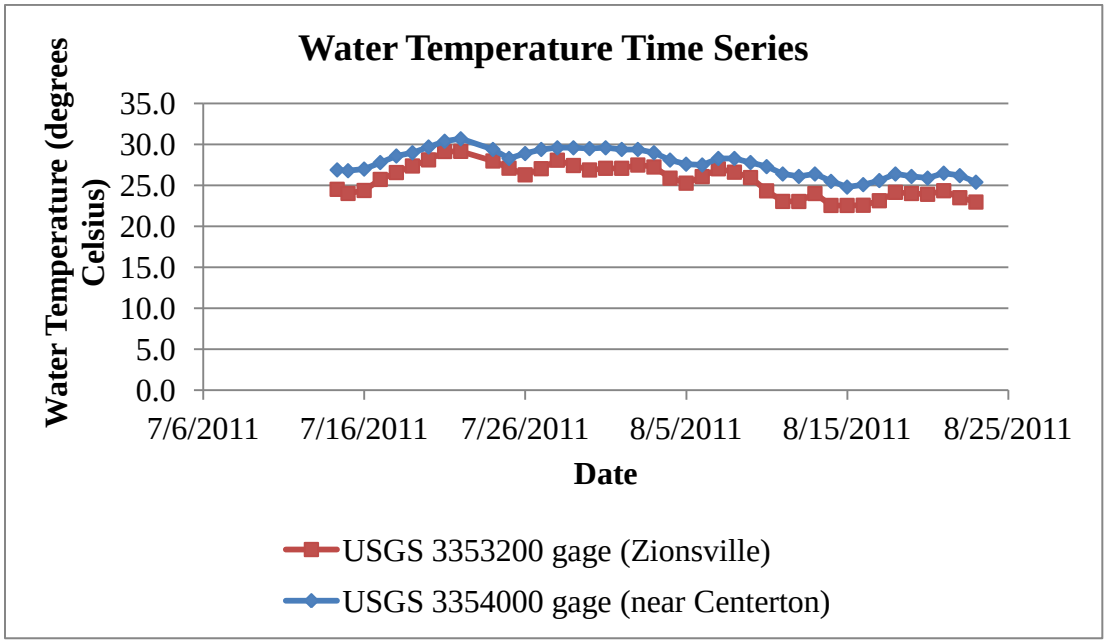


Figure 10: Water temperature time series comparison for USGS 3353200 and USGS gage 3354000.

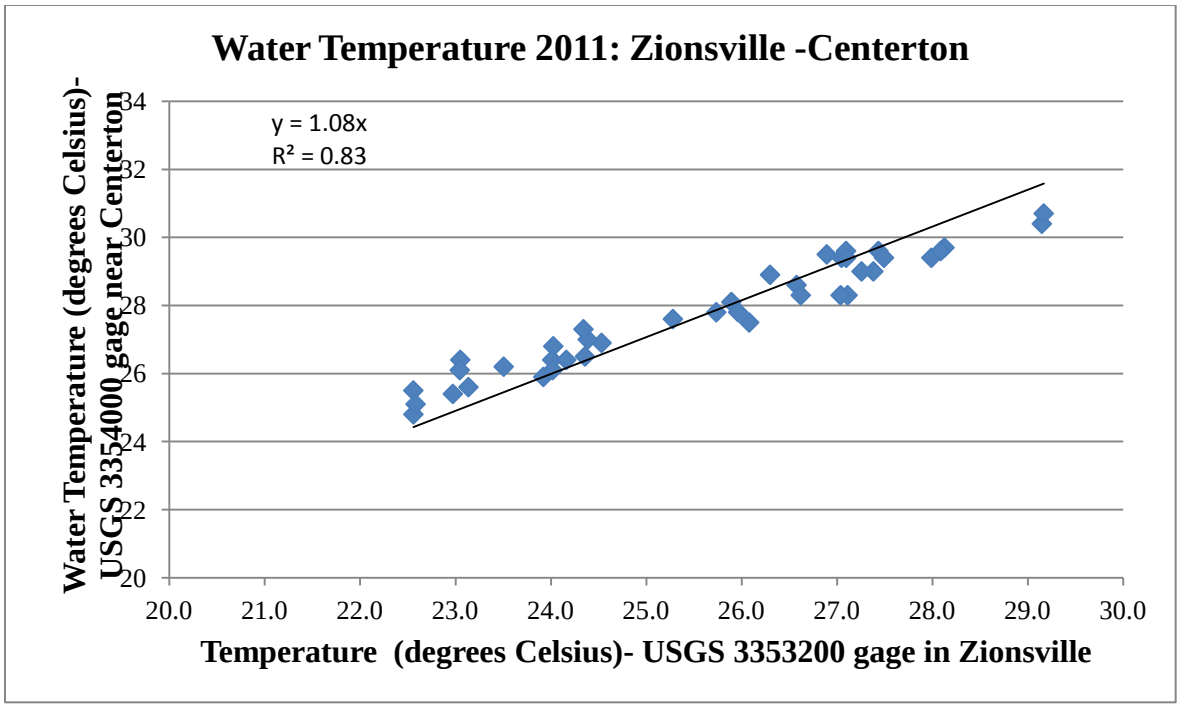


Figure 11: Water temperature regression
for USGS gage in Zionsville and near Centerton (IN).

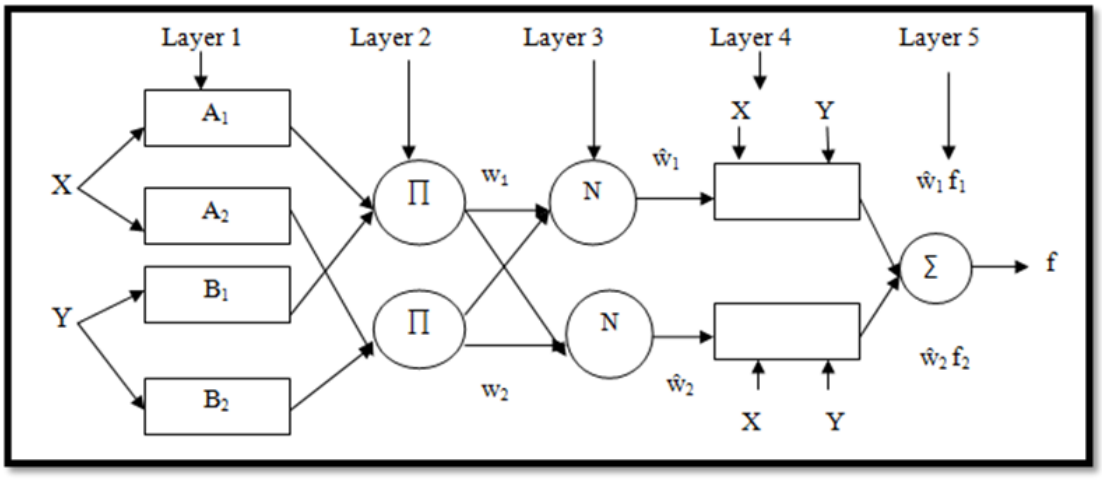


Figure 12: ANFIS Structure.

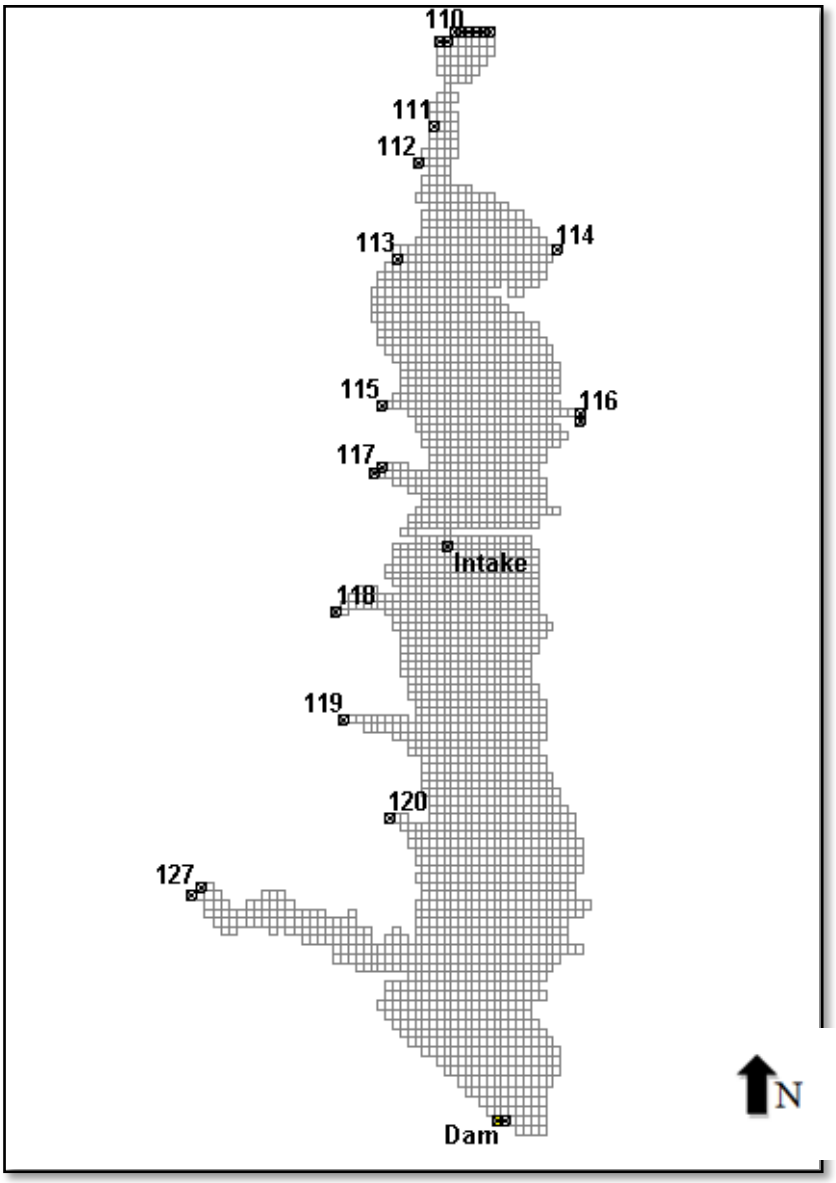


Figure 13: ECR grid cells assigned for boundary conditions.

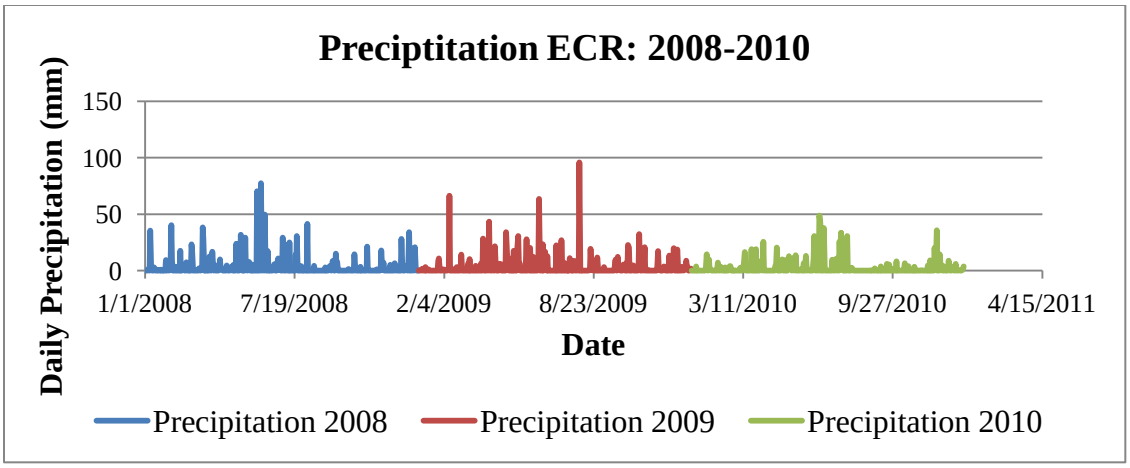


Figure 14: Daily precipitation for 2008-2010.

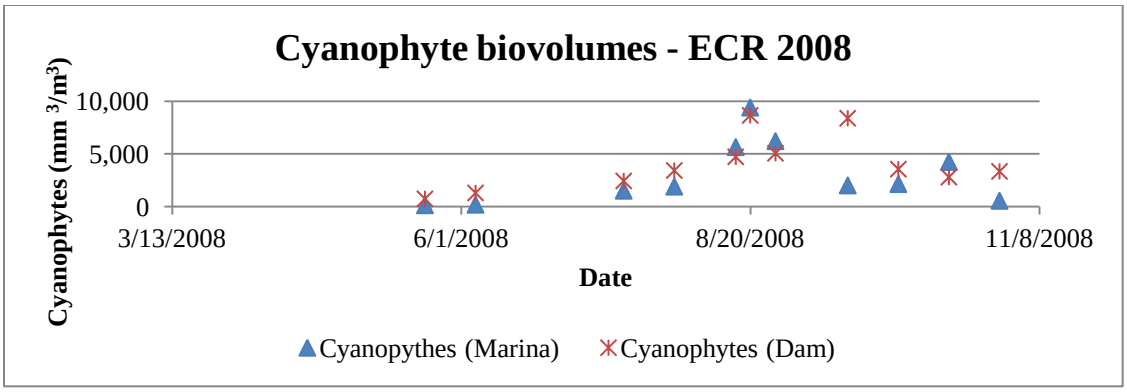


Figure 15: Cyanophyte biovolumes for Eagle Creek Reservoir sampled in marina and dam regions in 2008.

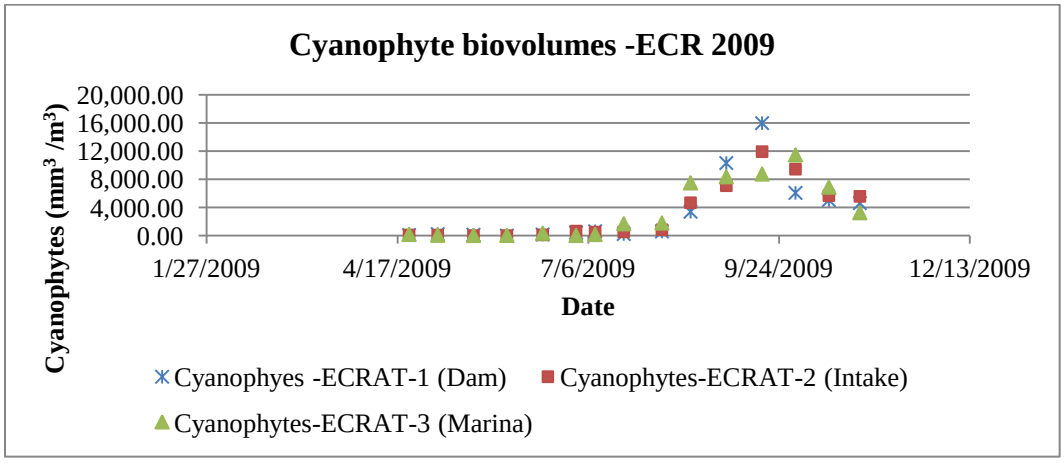


Figure 16: Cyanophyte biovolumes for Eagle Creek Reservoir sampled in marina, intake and dam regions in 2009.

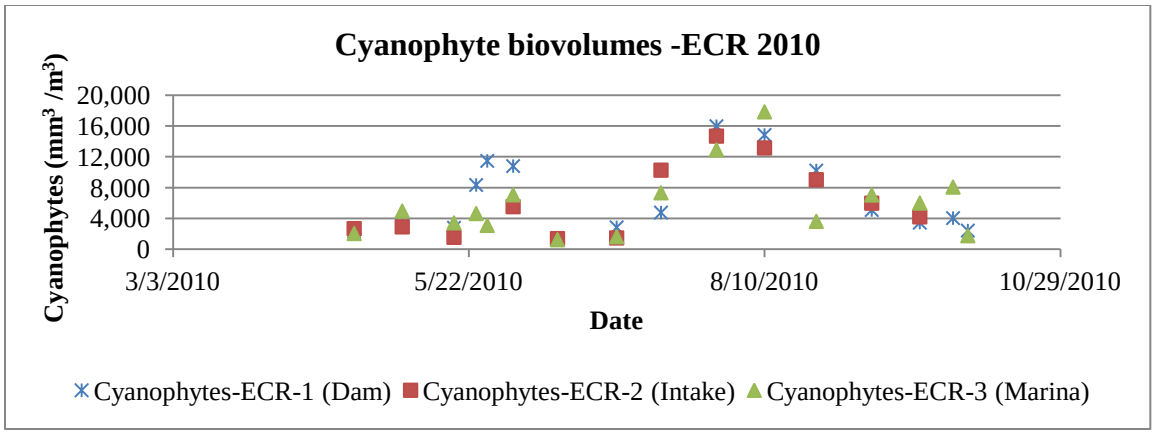


Figure 17: Cyanophyte biovolumes for Eagle Creek Reservoir sampled in marina, intake and dam regions in 2010.

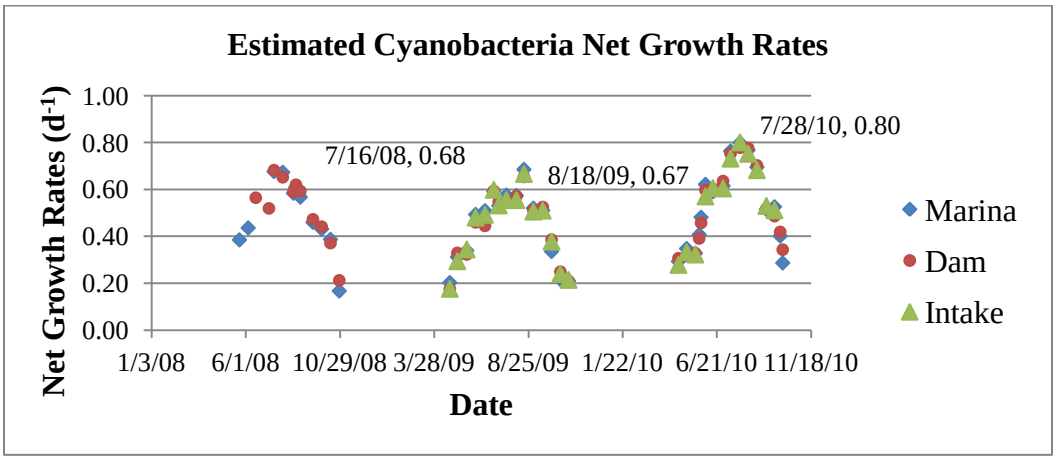


Figure 18: Estimated cyanobacteria net growth rates for three (3) regions in Eagle Creek Reservoir.
Growth rates were estimated based on equations described by Wong et al. (2009).

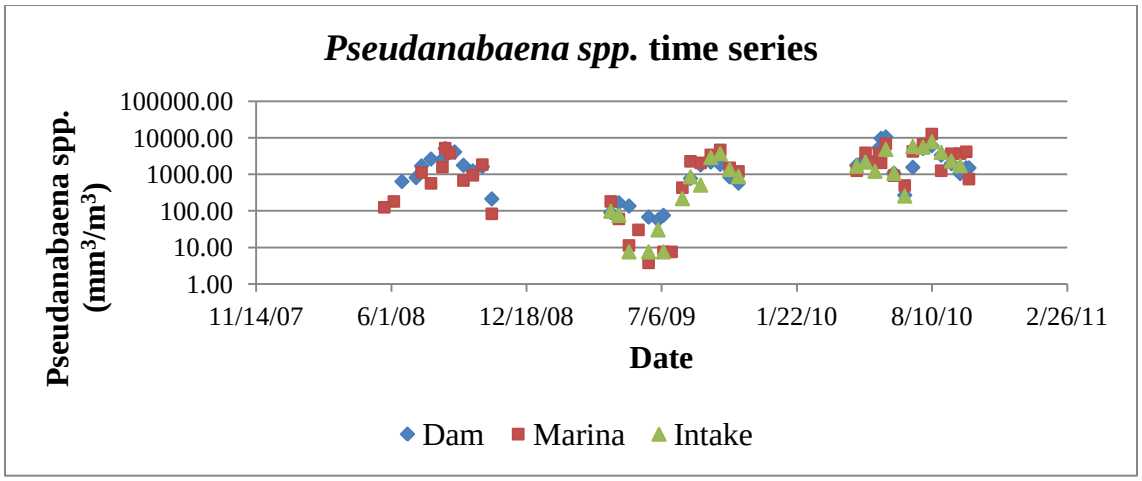


Figure 19: *Pseudanabaena spp*. biovolumes for Eagle Creek Reservoir (2008-2010).

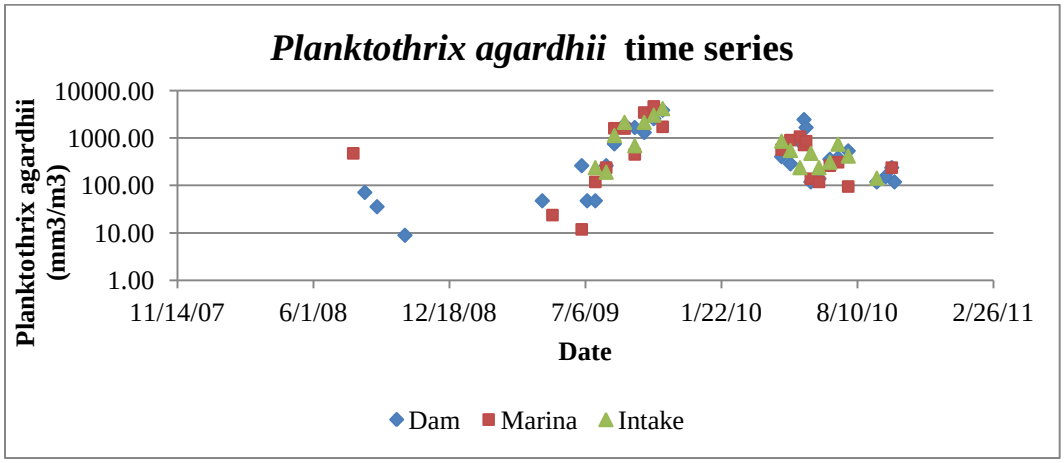


Figure 20: *Planktothrix agardhii* biovolumes for Eagle Creek Reservoir in 2008-2010.

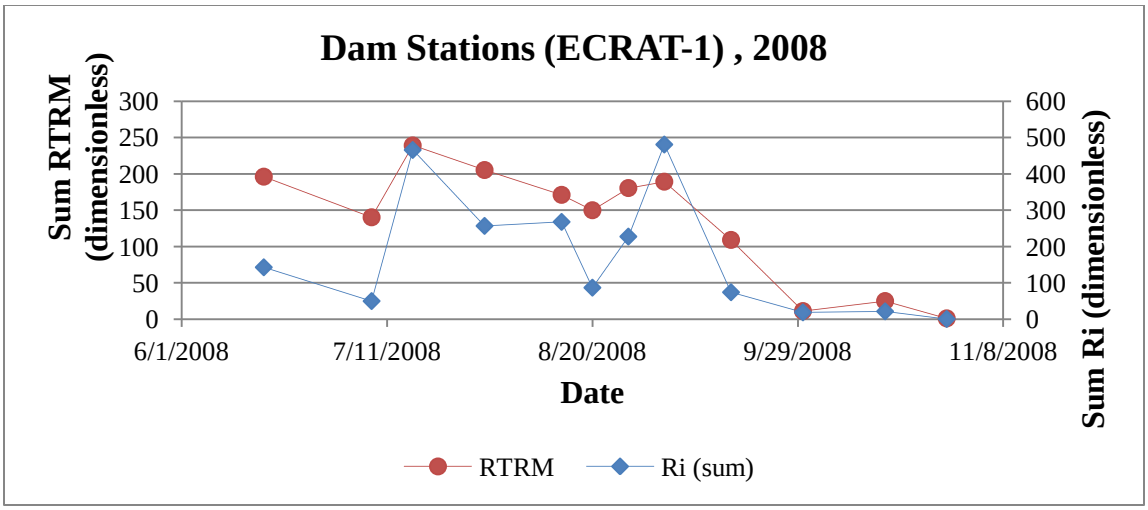


Figure 21: Sum RTRM and R_i averaged values for Dam Region in 2008.

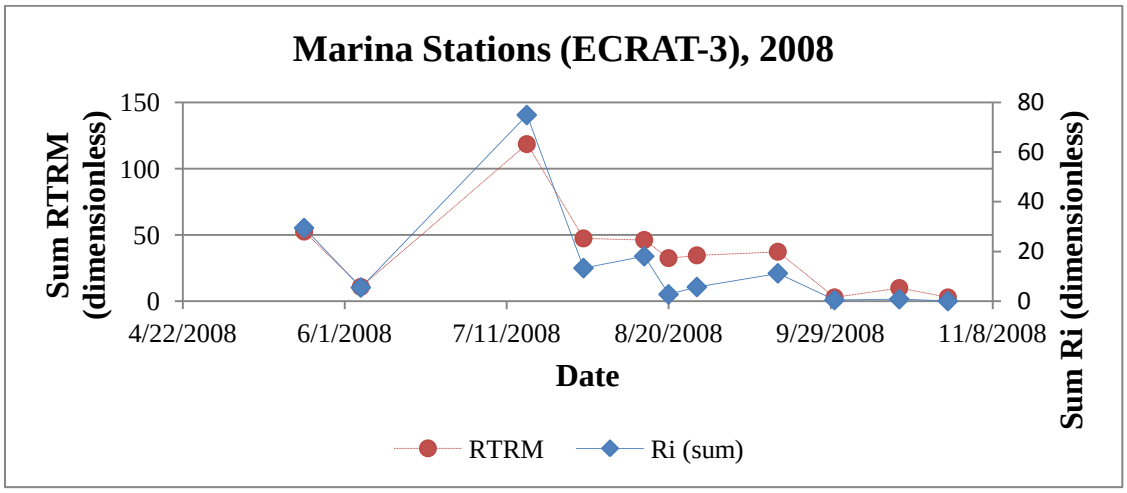


Figure 22: Sum RTRM and R_i averaged values for Marina Region in 2008.

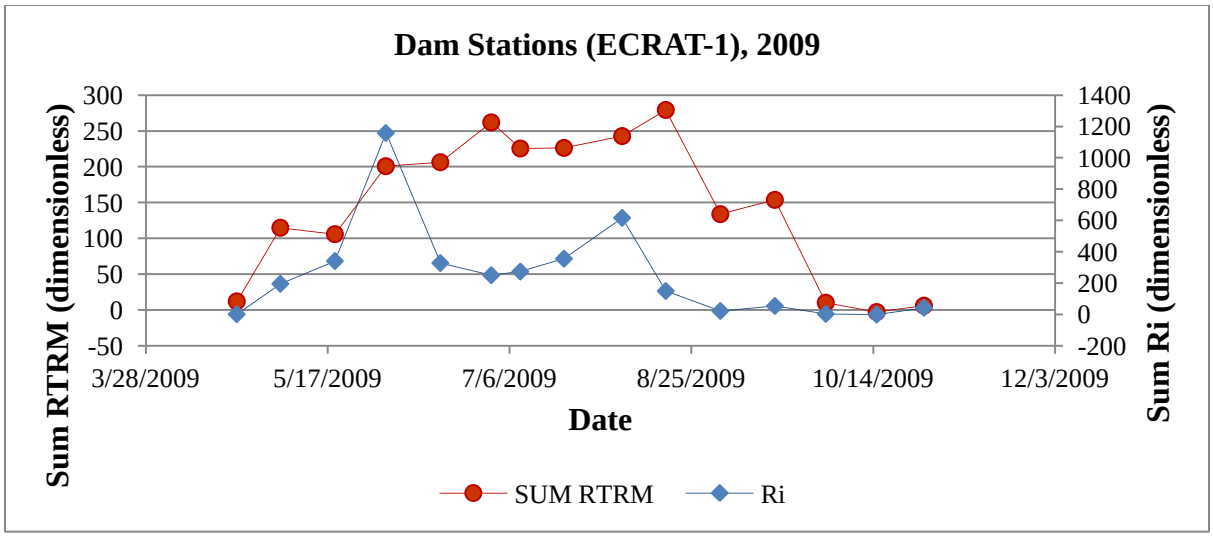


Figure 23: Sum RTRM and R_i averaged values for Dam Region in 2009.

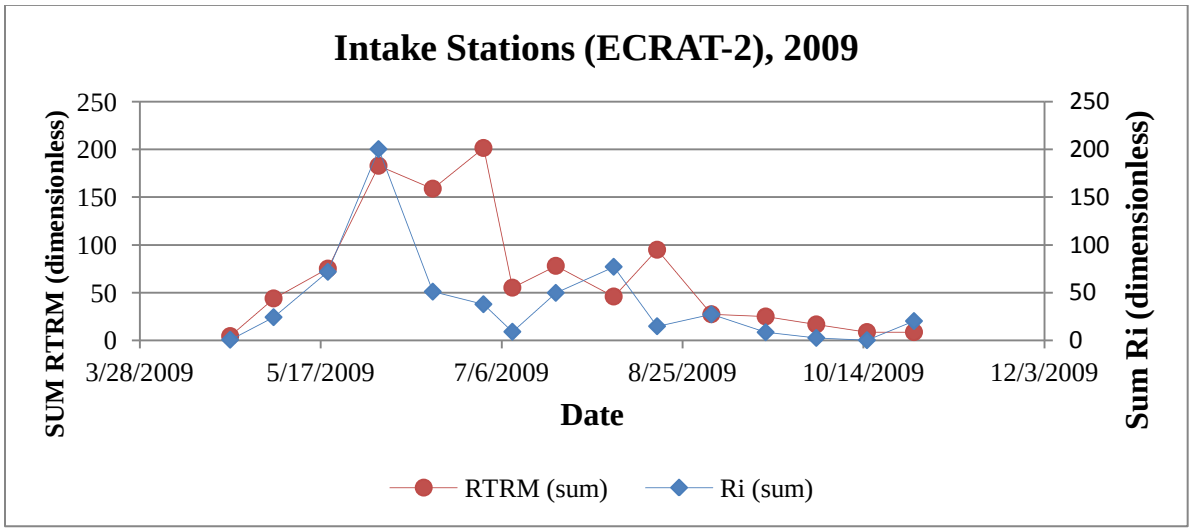


Figure 24: Sum RTRM and R_i averaged values for Intake Region in 2009.

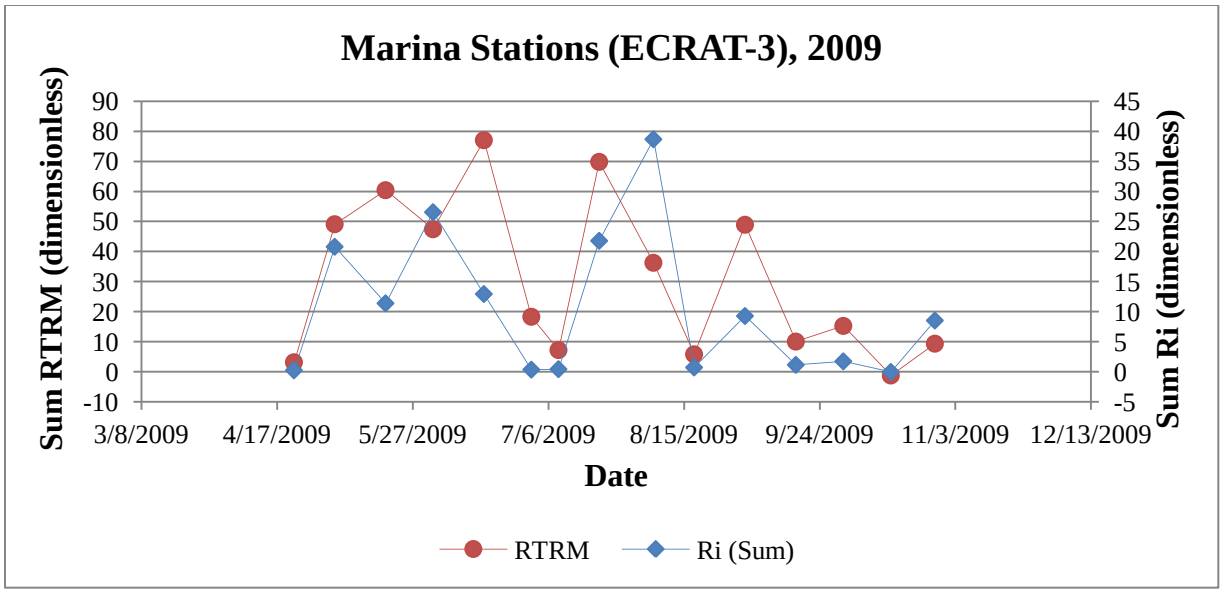


Figure 25: Sum RTRM and R_i averaged values for Marina Region in 2009.

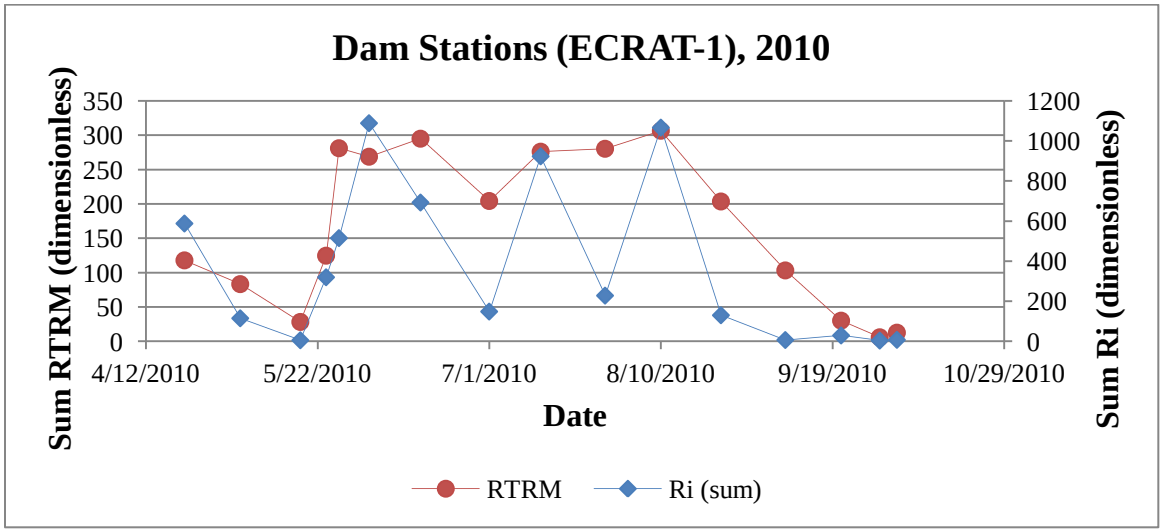


Figure 26: Sum RTRM and R_i averaged values for Dam Region in 2010.

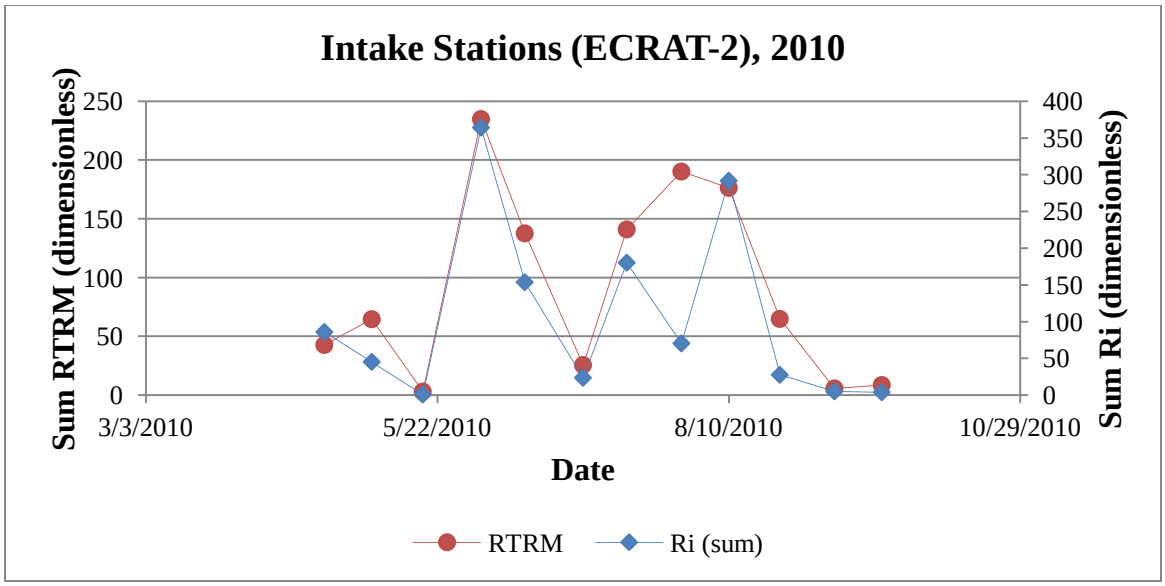


Figure 27: Sum RTRM and R_i averaged values for Intake Region in 2010.

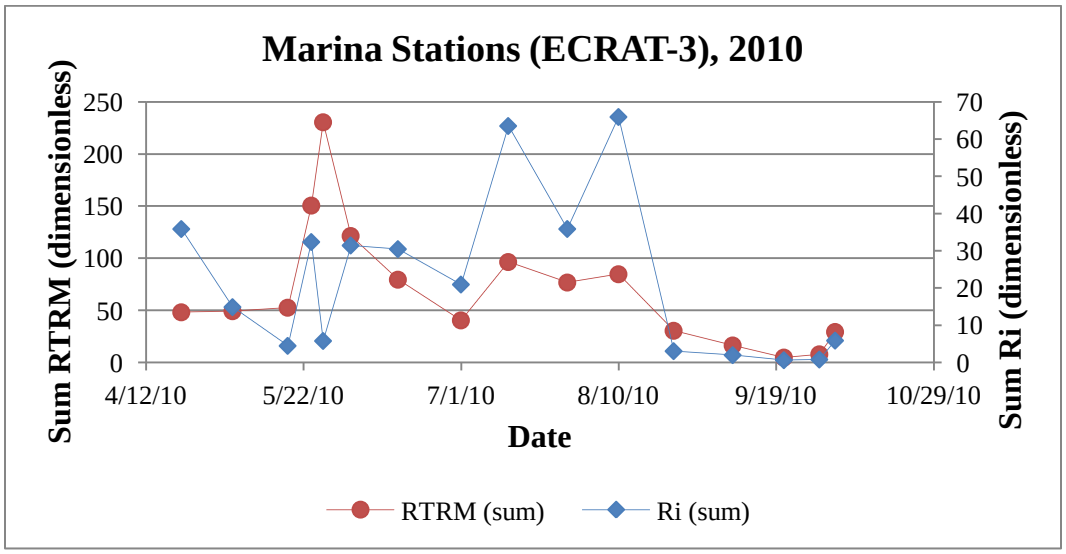


Figure 28: Sum RTRM and R_i averaged values for Marina Region in 2010.

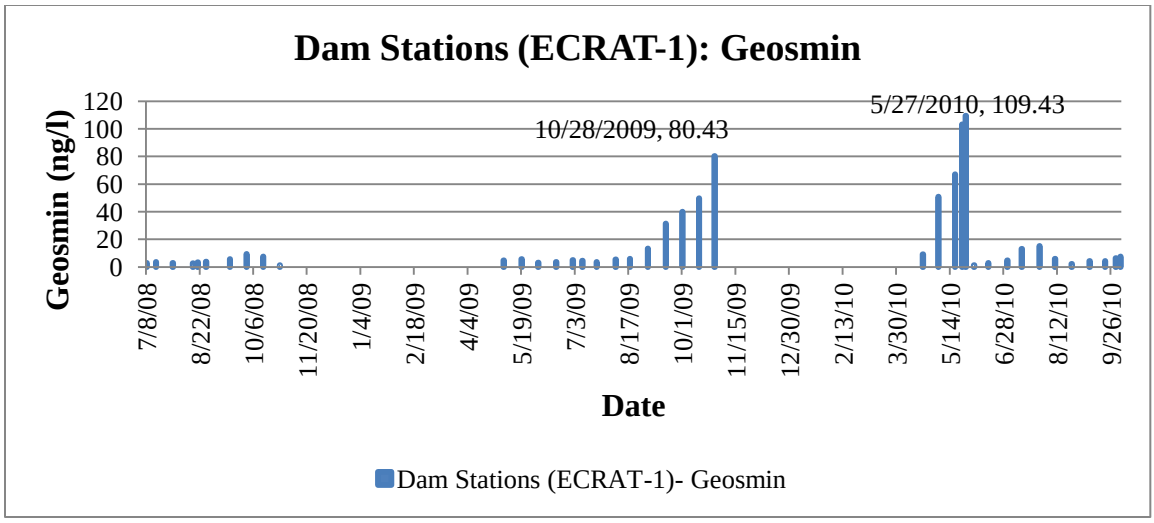


Figure 29: Geosmin concentrations for Dam Region (2008-2010).

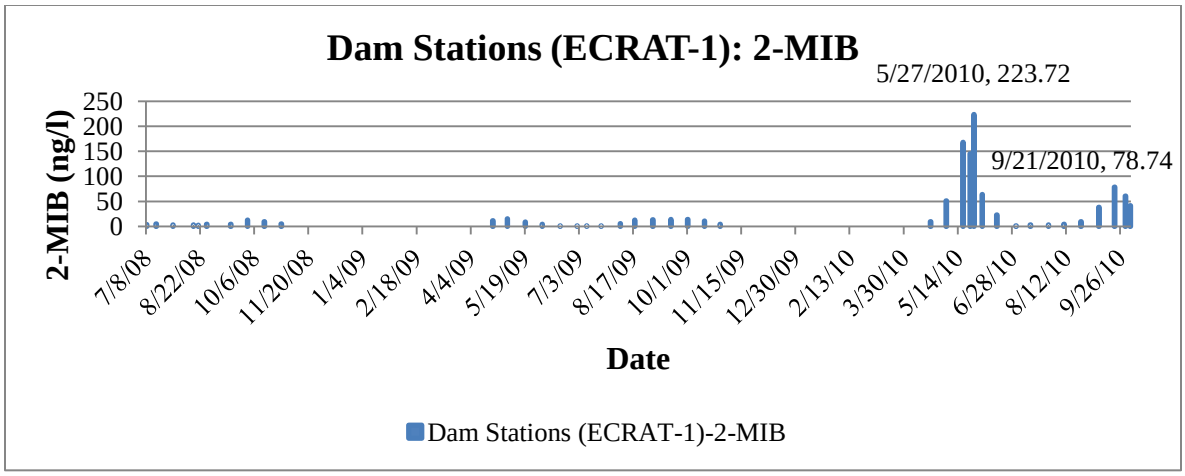


Figure 30: 2-MIB concentrations for Dam Region (2008-2010).

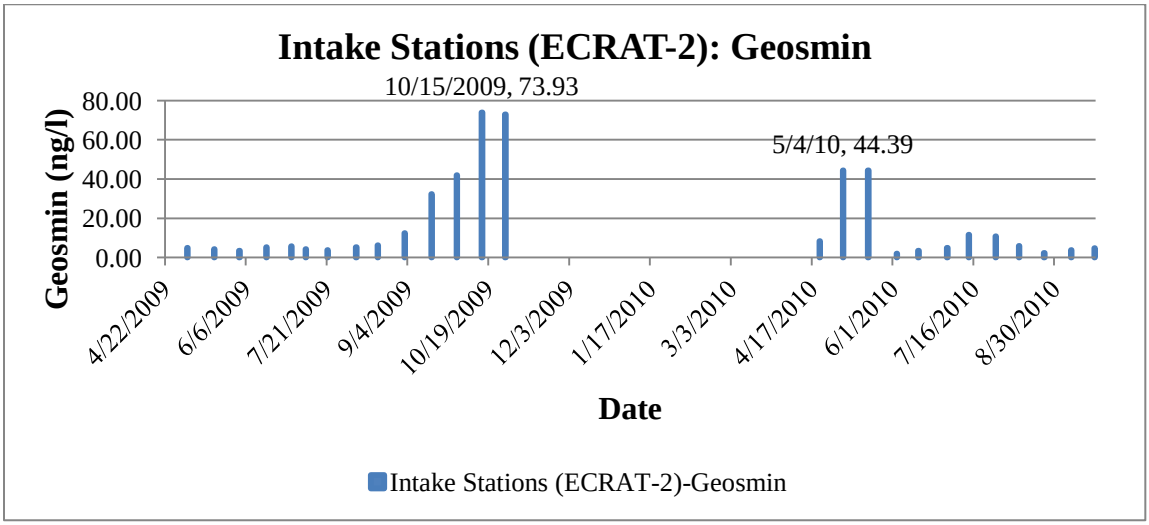


Figure 31: Geosmin concentrations for Intake Region (2009-2010).

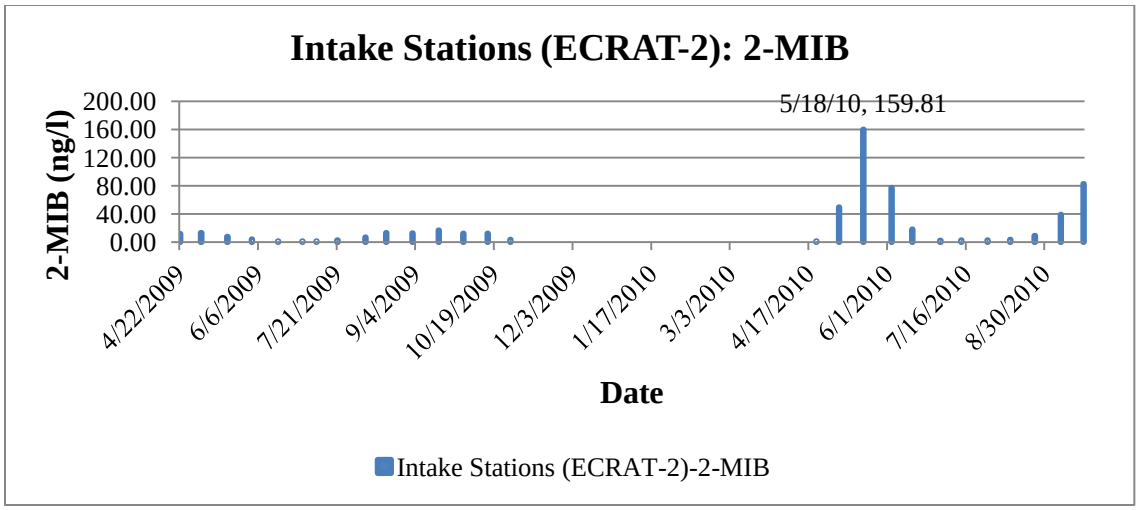


Figure 32: 2-MIB concentrations for Intake Region (2009-2010).

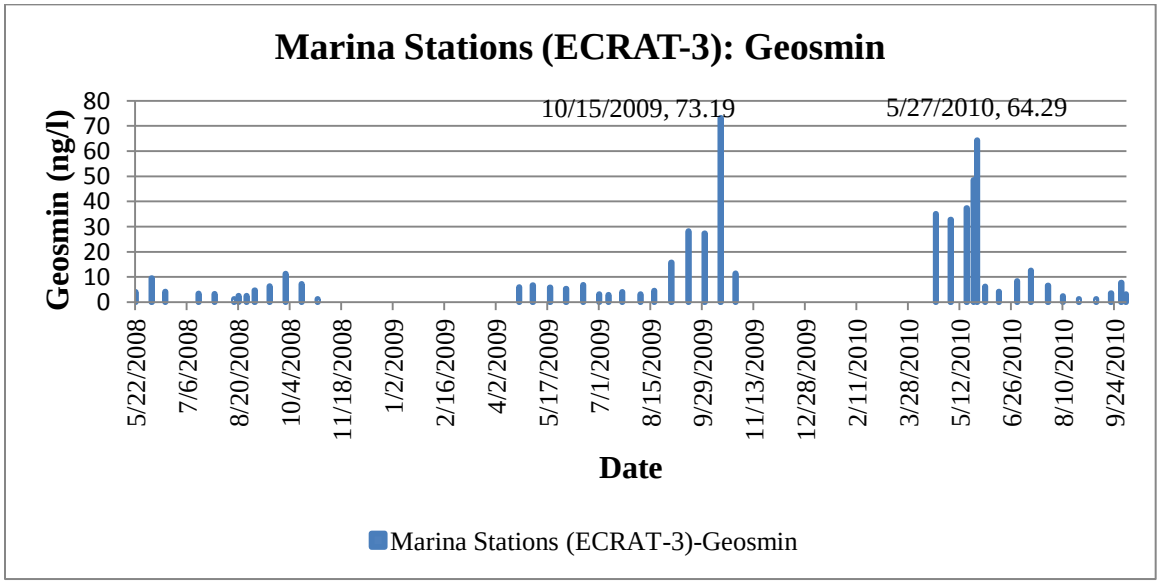


Figure 33: Geosmin concentrations for Marina Region (2008-2010).

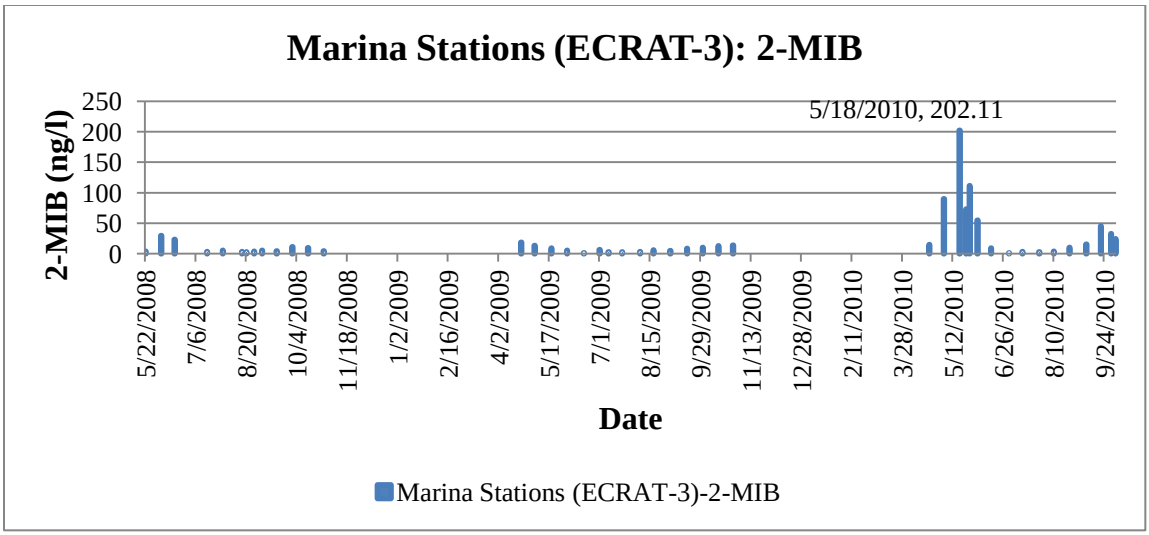


Figure 34: 2-MIB concentrations for Marina Region (2008-2010).

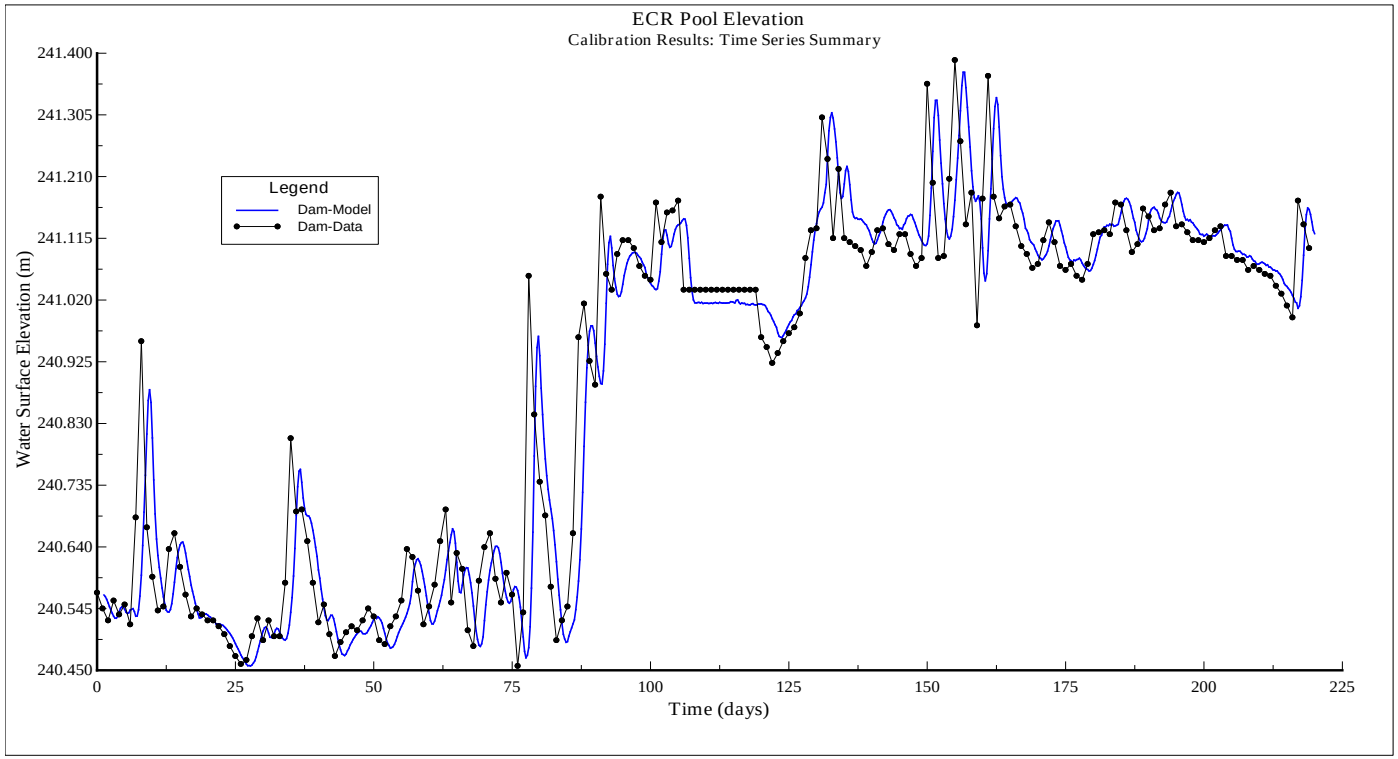


Figure 35: Pool elevation calibration results for ECR (2008).

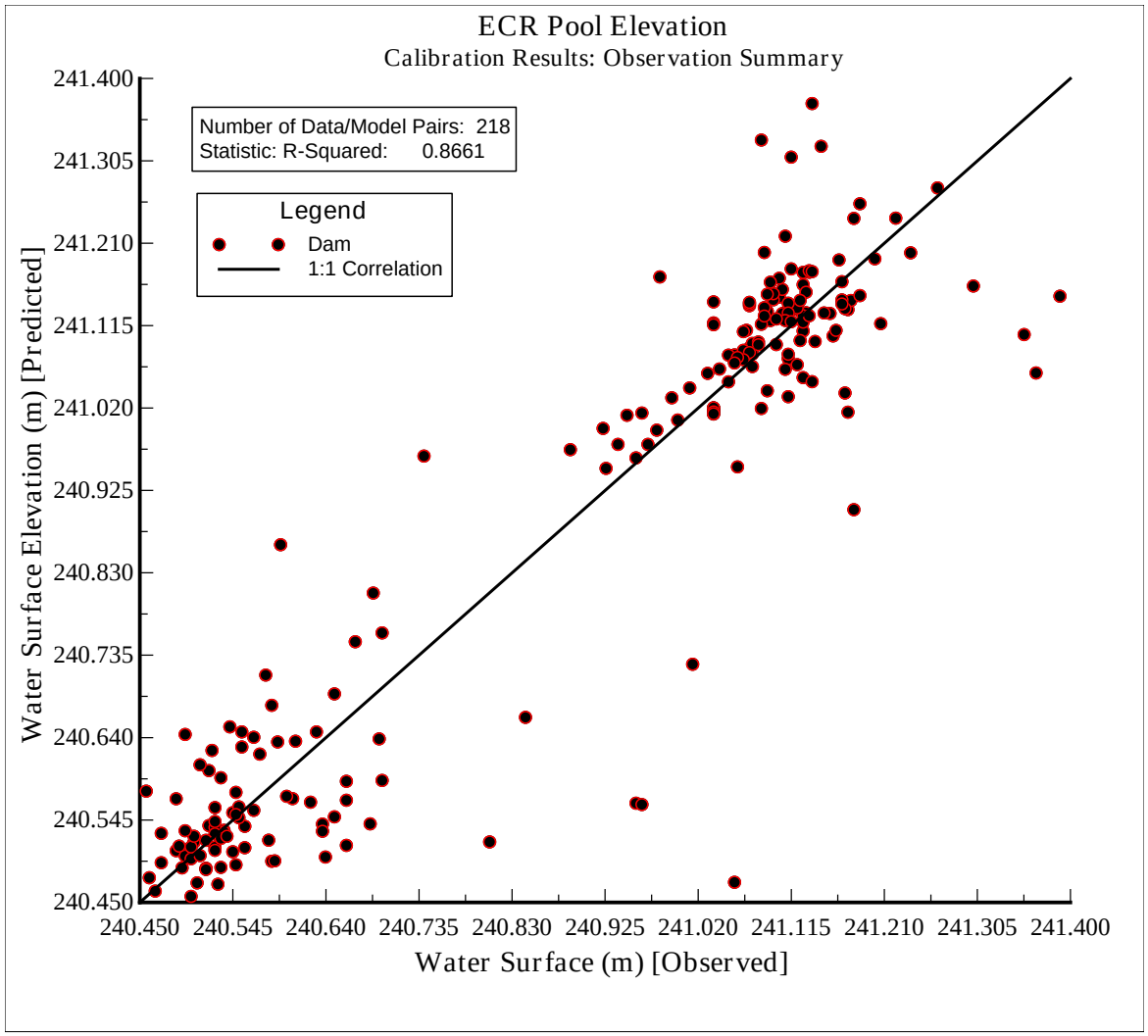


Figure 36: Coefficient of determination results for calibrated water surface elevation for ECR for 2008.

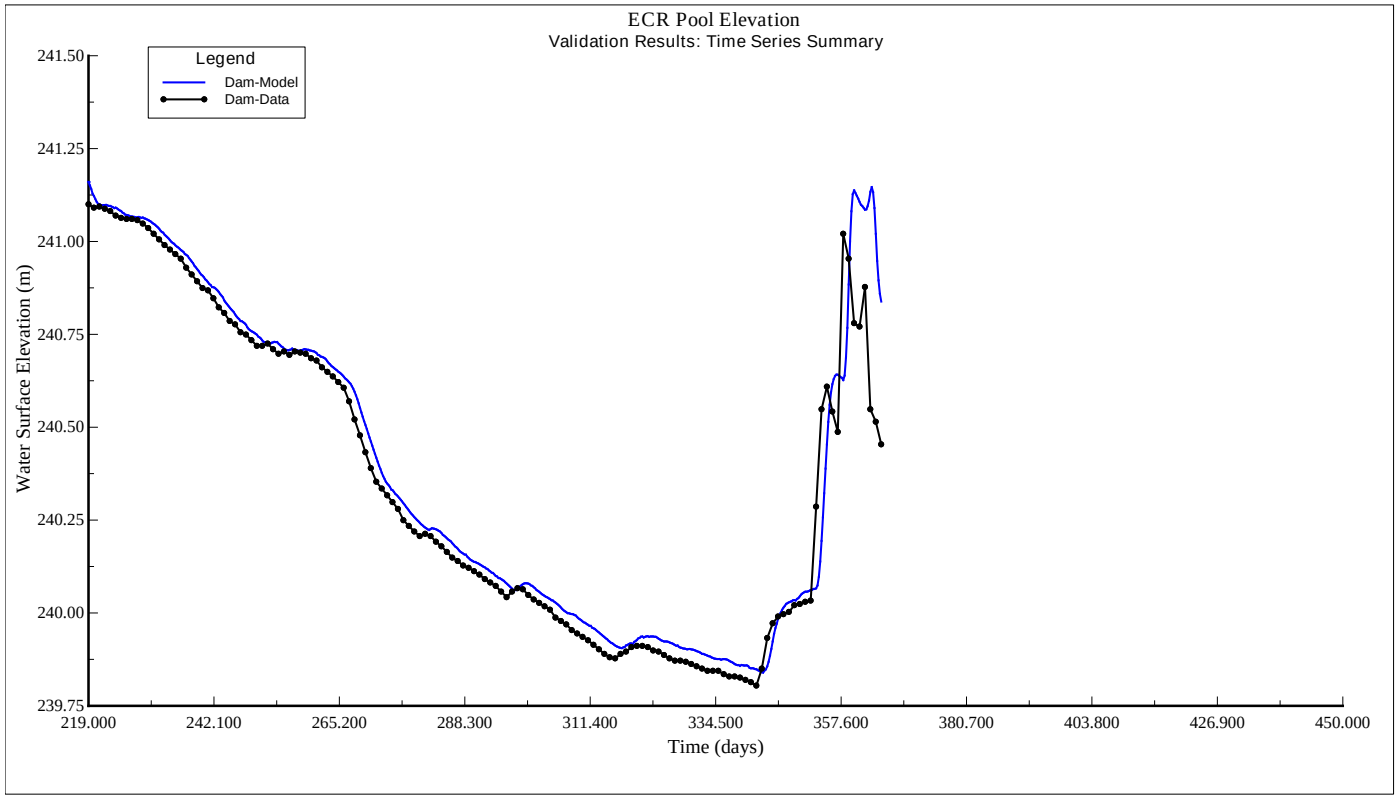


Figure 37: Pool elevation validation results for ECR (2008).

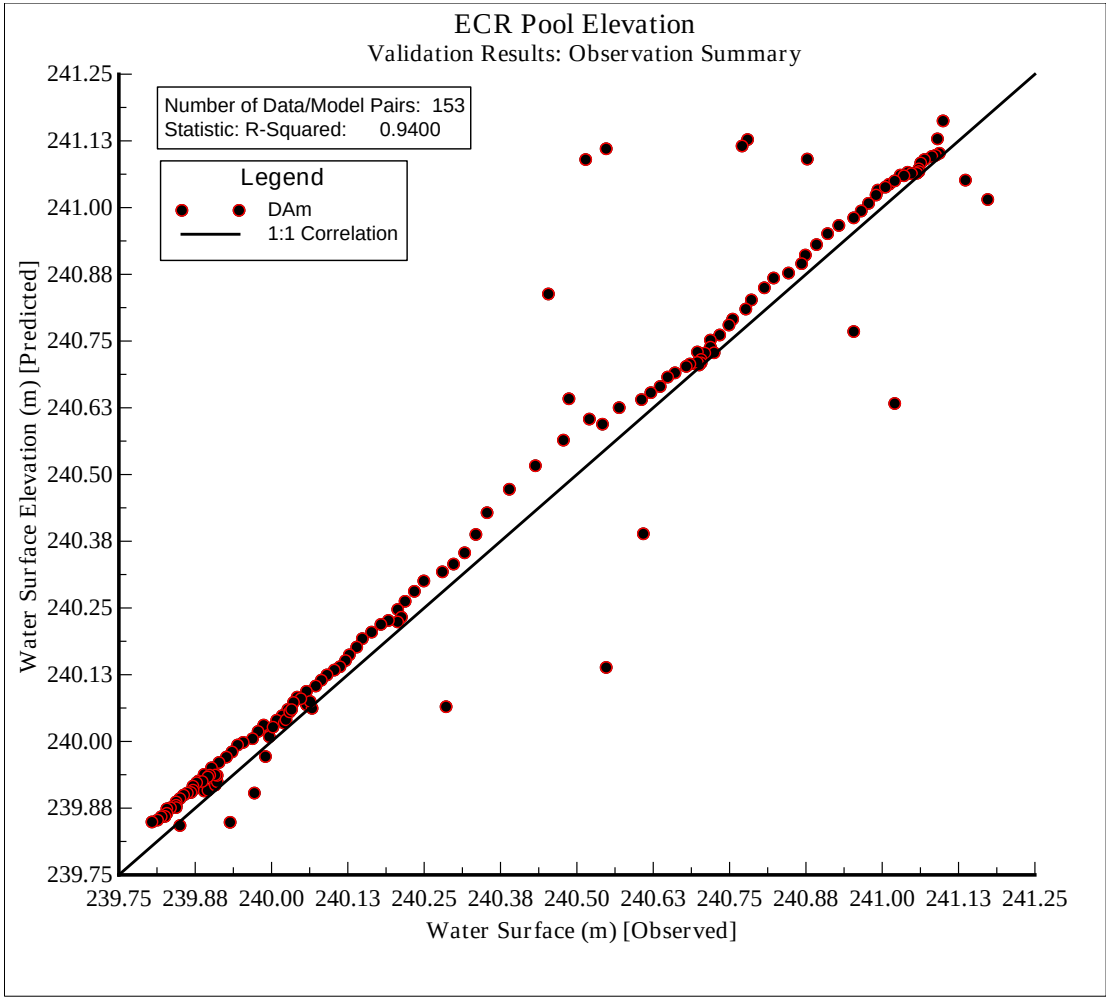
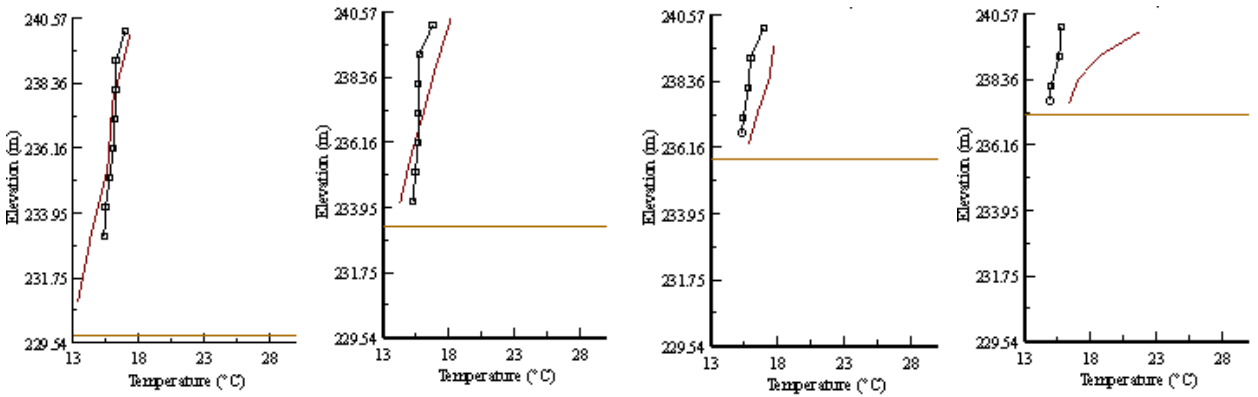


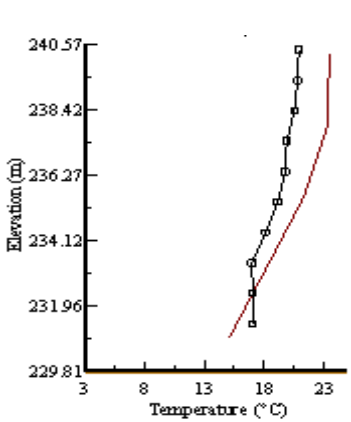
Figure 38: Coefficient of determination results

for validated water surface elevation for ECR for 2008.

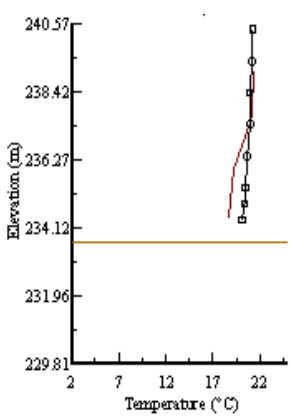
Figure 39 a-u: Temperature calibration vertical profiles for monitoring stations A1-F3. Simulated values represented as solid line and measured as line with squares/circles.



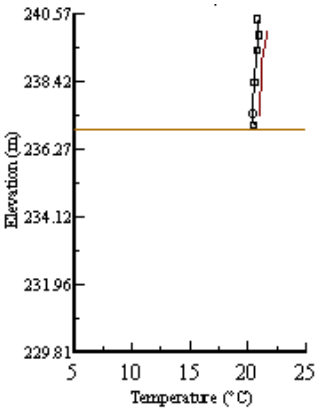
a. A1 b. A2 c. A3 d. A4



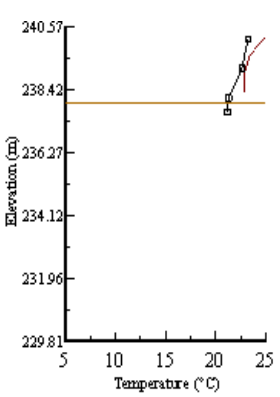
e. B1



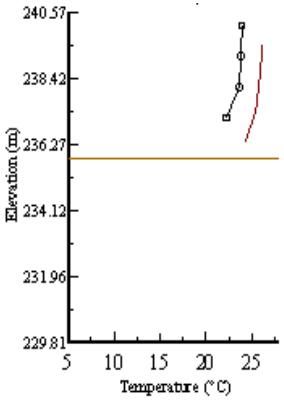
f. B2



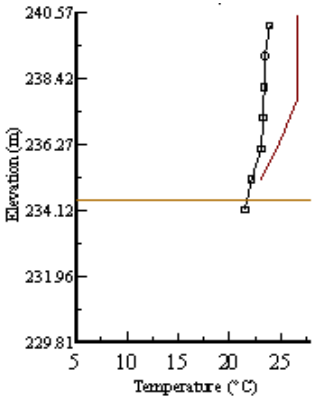
g. B3



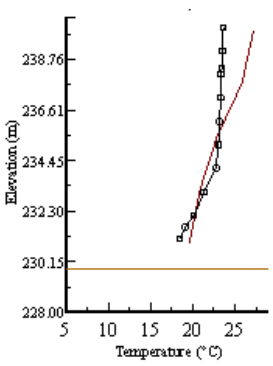
h. C1



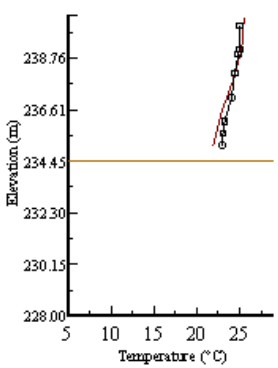
i. C2



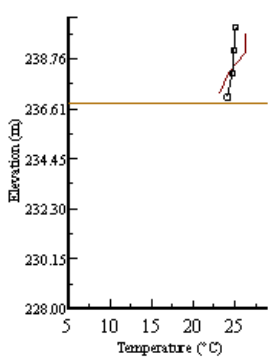
j. C3



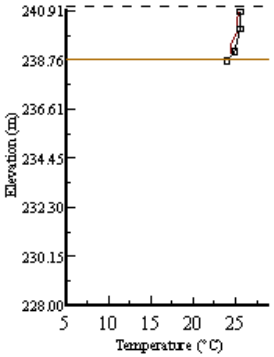
k. D1



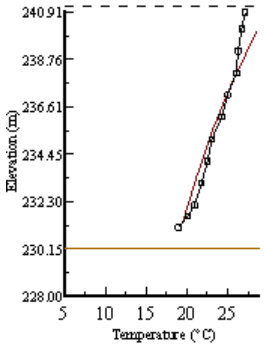
l. D2



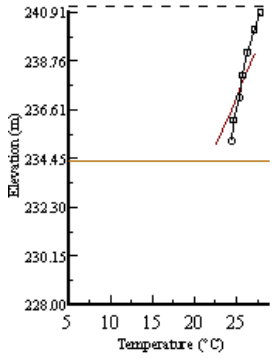
m. D3



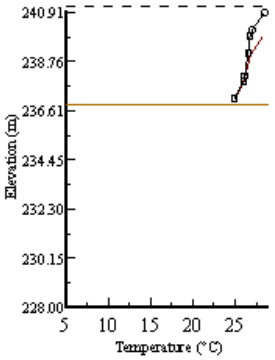
n. D4



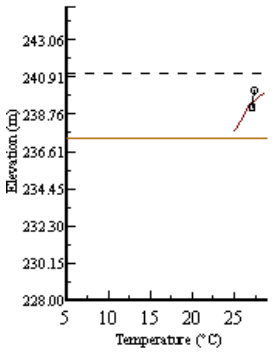
o. E1



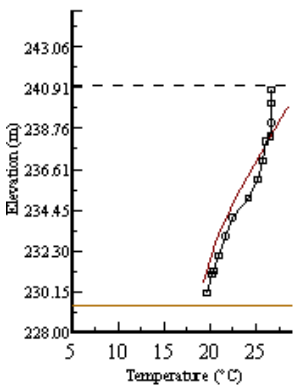
p. E2



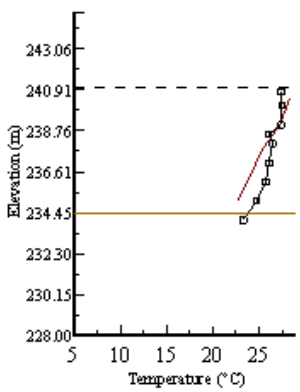
q. E3



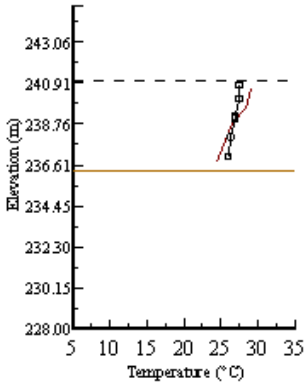
r. E4


243.06
240.91
238.76
236.61
234.45
232.30
230.15
228.00
Elevation (m)
5 10 15 20 25
Temperature (°C)

s. F1

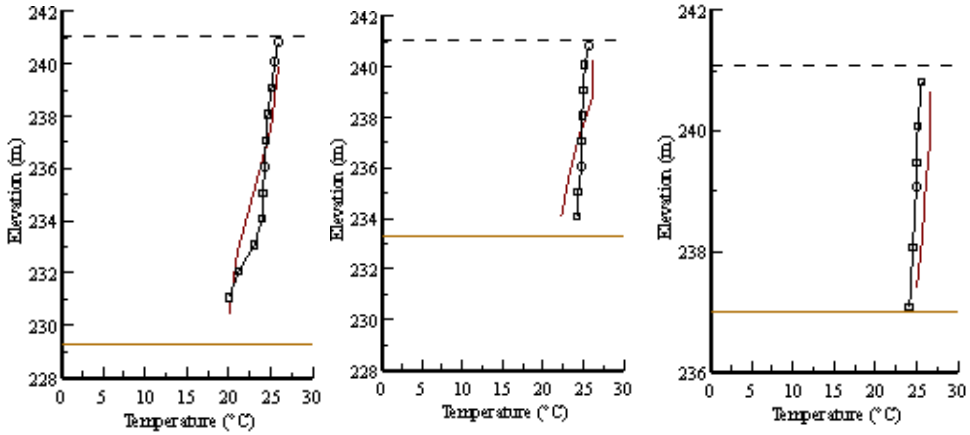

243.06
240.91
238.76
236.61
234.45
232.30
230.15
228.00
Elevation (m)
5 10 15 20 25
Temperature (°C)

t. F2

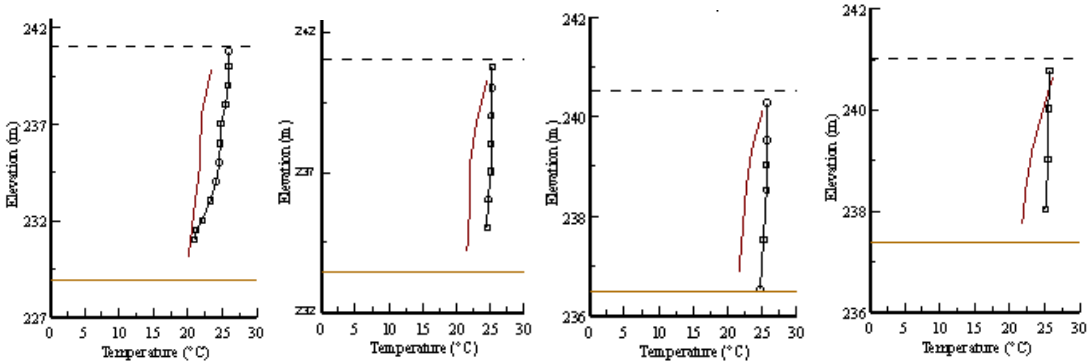

243.06
240.91
238.76
236.61
234.45
232.30
230.15
228.00
Elevation (m)
5 10 15 20 25 30 35
Temperature (°C)

u. F3

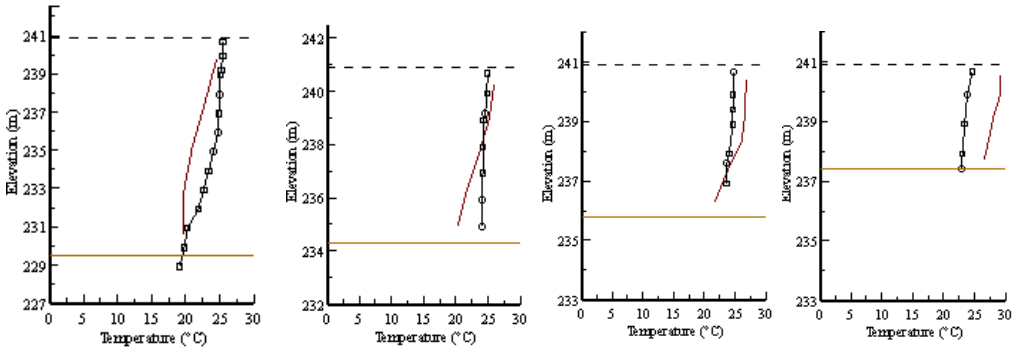
Figure 40 a-zc: Temperature validation vertical profiles for monitoring stations G1-N4. Simulated values represented as solid line and measured as line with squares/circles



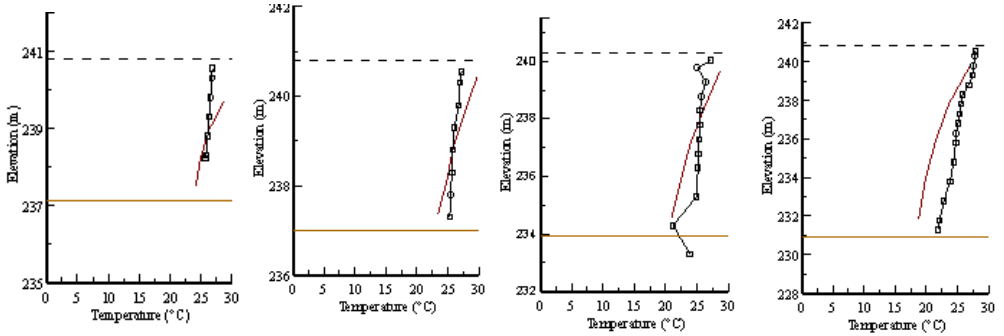
a. Station G1 b. Station G2 c. Station G3



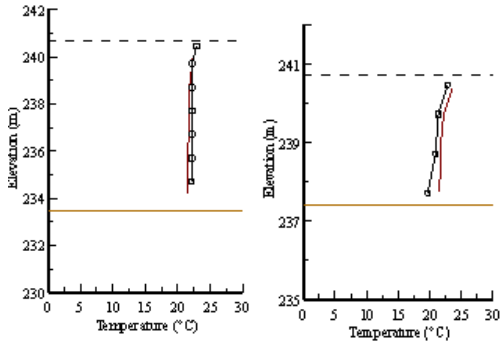
d. Station H1 e. Station H2 f. Station H3 g. Station H4



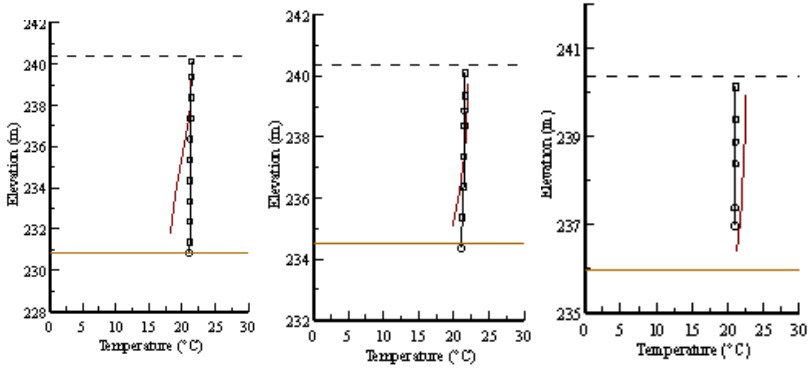
h. Station I1 i. Station I2 j. Station I3 k. Station I4



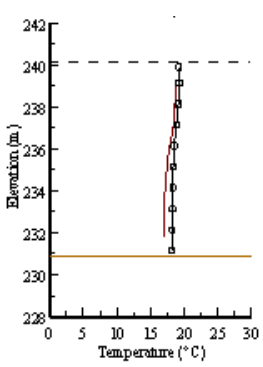
l. Station J1 m. Station J2 n. Station J3 o. Station J4



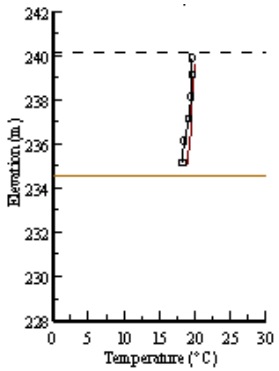
o. Station K2 r. Station K4



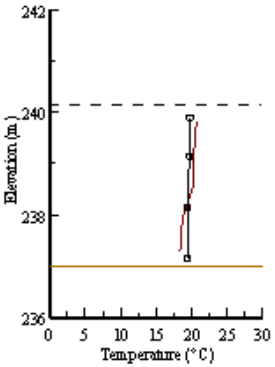
s.L1 t. L2 u.L3



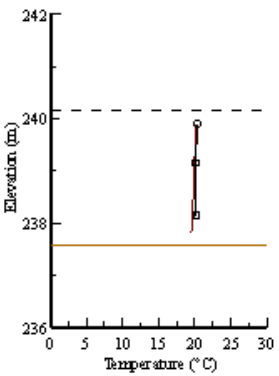
v. Station M1



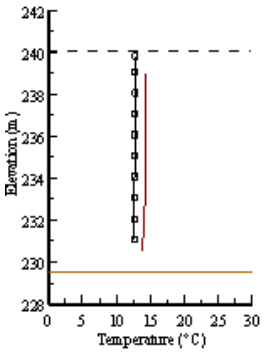
w. Station M2



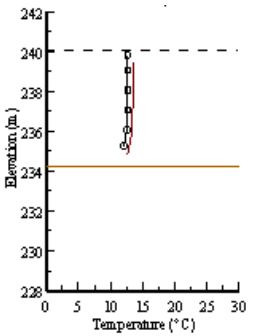
x. Station M3



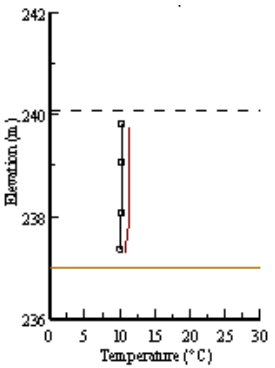
y. Station M4



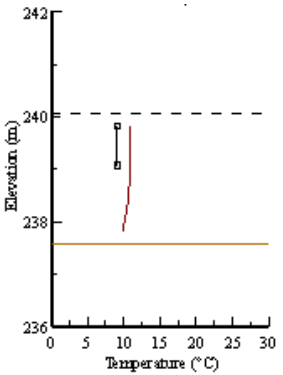
z. Station N1



za. Station N2



zb. Station N3



zc. Station N4

Figure 41 a: EFDC water quality calibration results

for total phosphorus for Dam monitoring station.

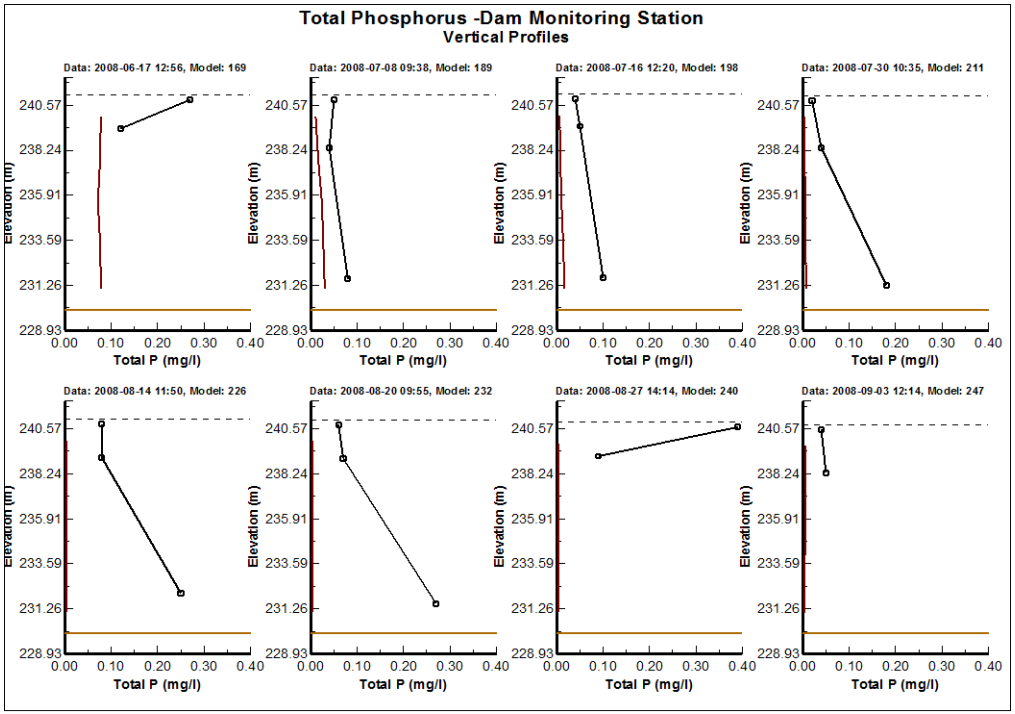


Figure 41 b: EFDC water quality calibration results for total phosphorus for Marina monitoring station

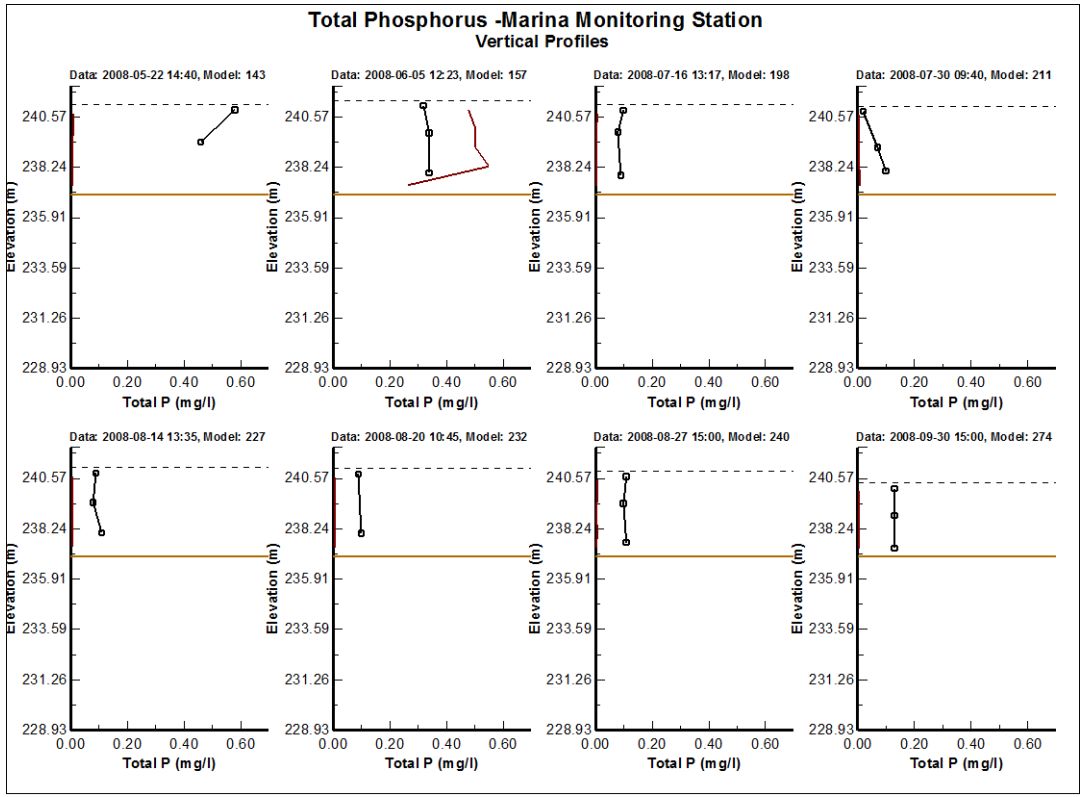


Figure 42 a: EFDC water quality calibration results

for ammonia nitrogen for Dam monitoring station.

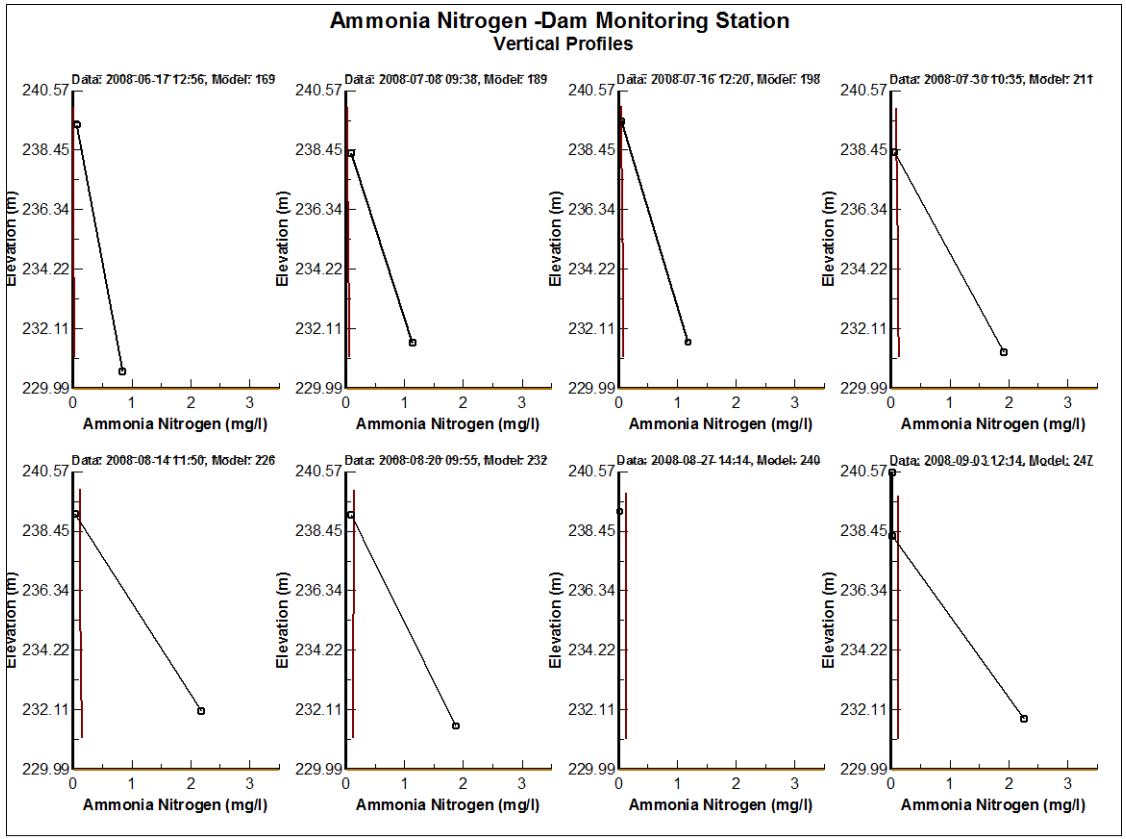


Figure 42 b: EFDC water quality calibration results

for ammonia nitrogen for Marina monitoring station.

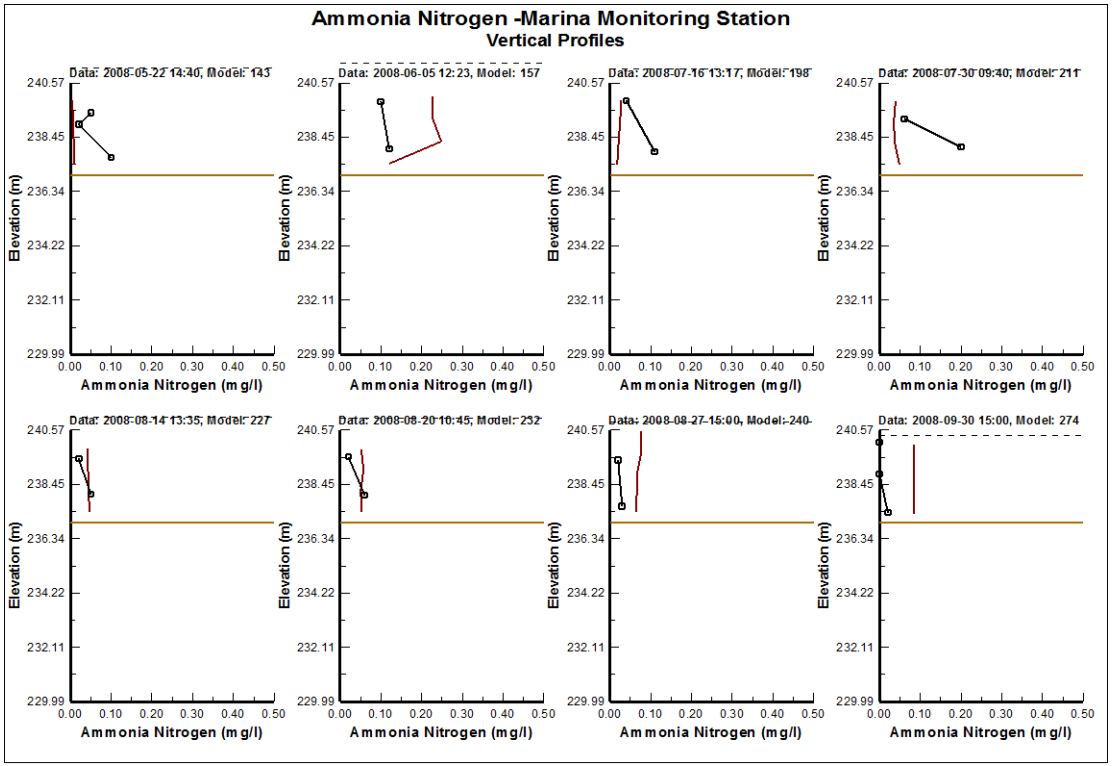


Figure 43 a: EFDC water quality calibration results

for nitrate-nitrogen for Dam monitoring station.

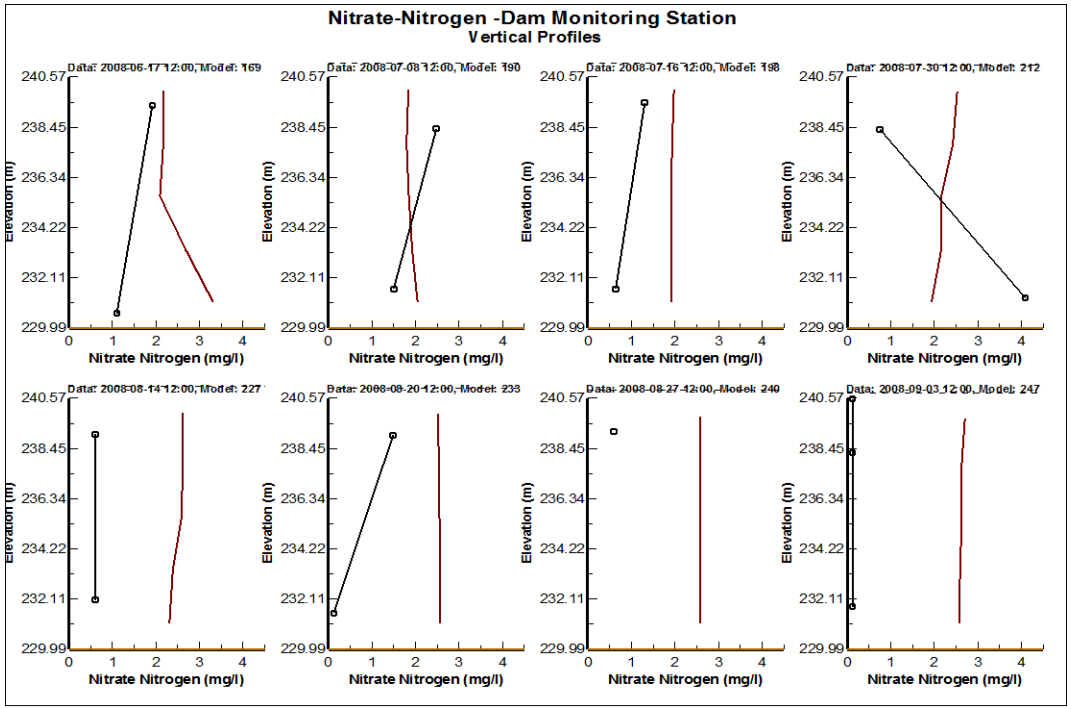


Figure 43 b: EFDC water quality calibration results

for nitrate-nitrogen for Marina monitoring station.

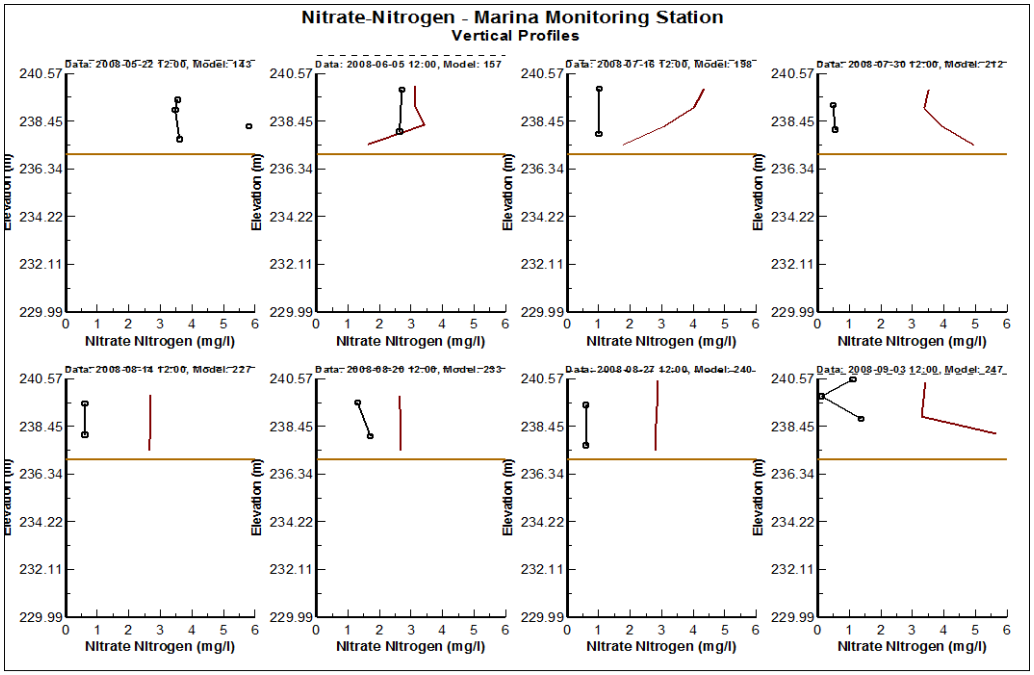


Figure 44 a: EFDC water quality calibration results

for dissolved oxygen for Dam monitoring station.

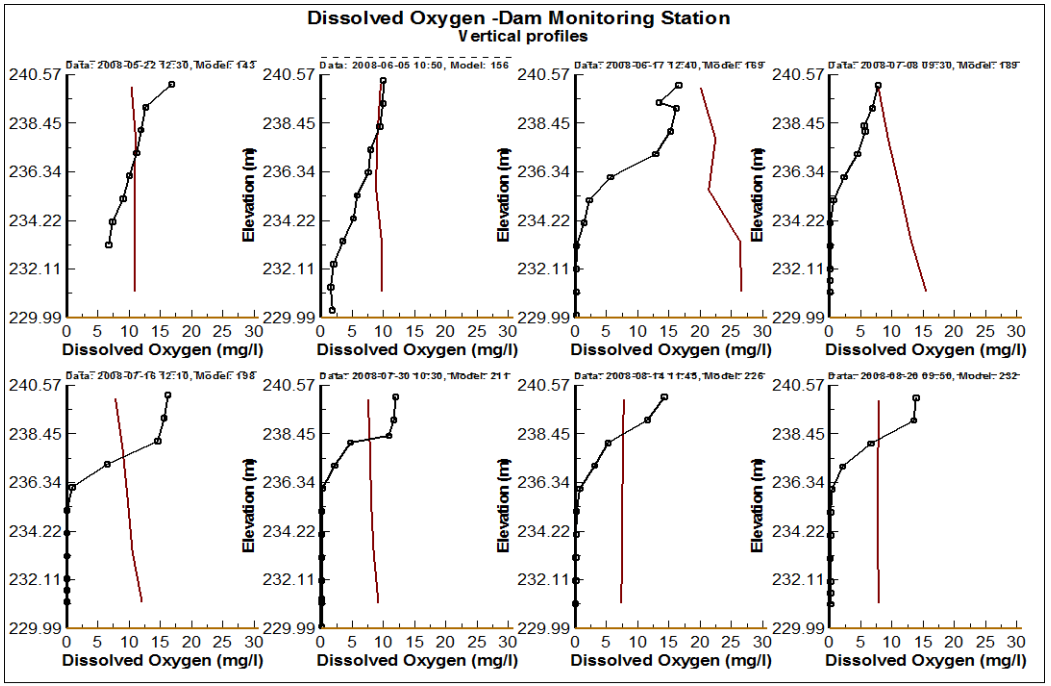


Figure 44 b: EFDC water quality calibration results for dissolved oxygen for Marina monitoring station.

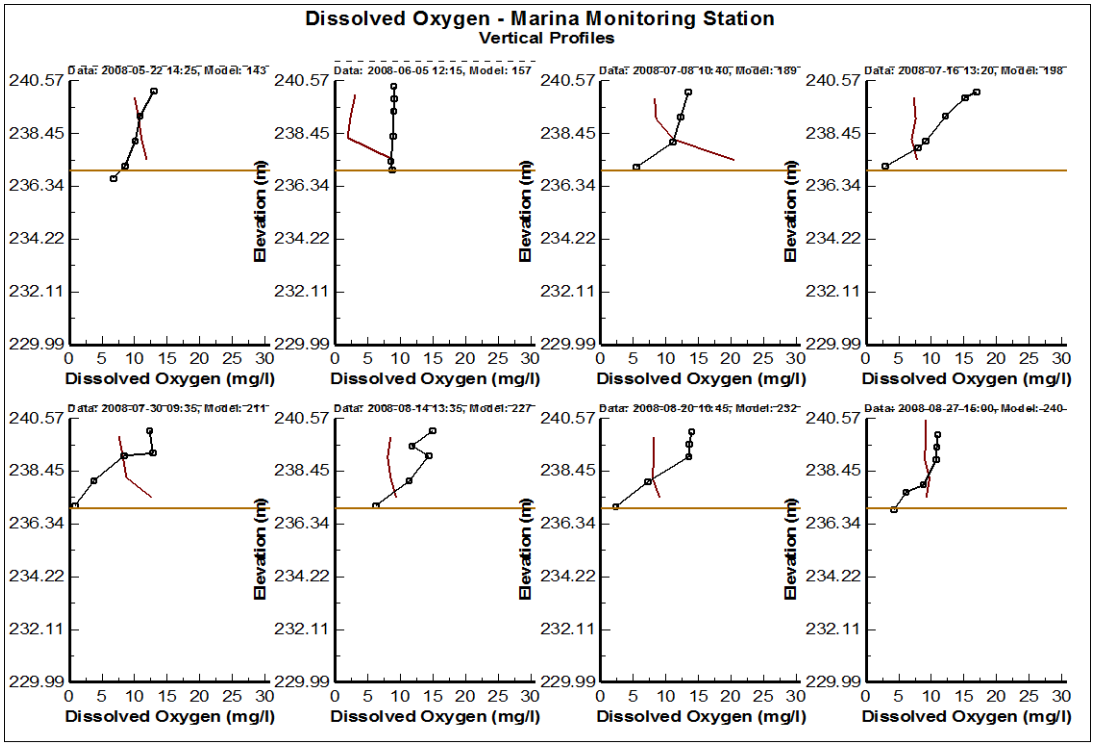


Figure 44 c: EFDC water quality calibration results for dissolved oxygen for Intake monitoring station.

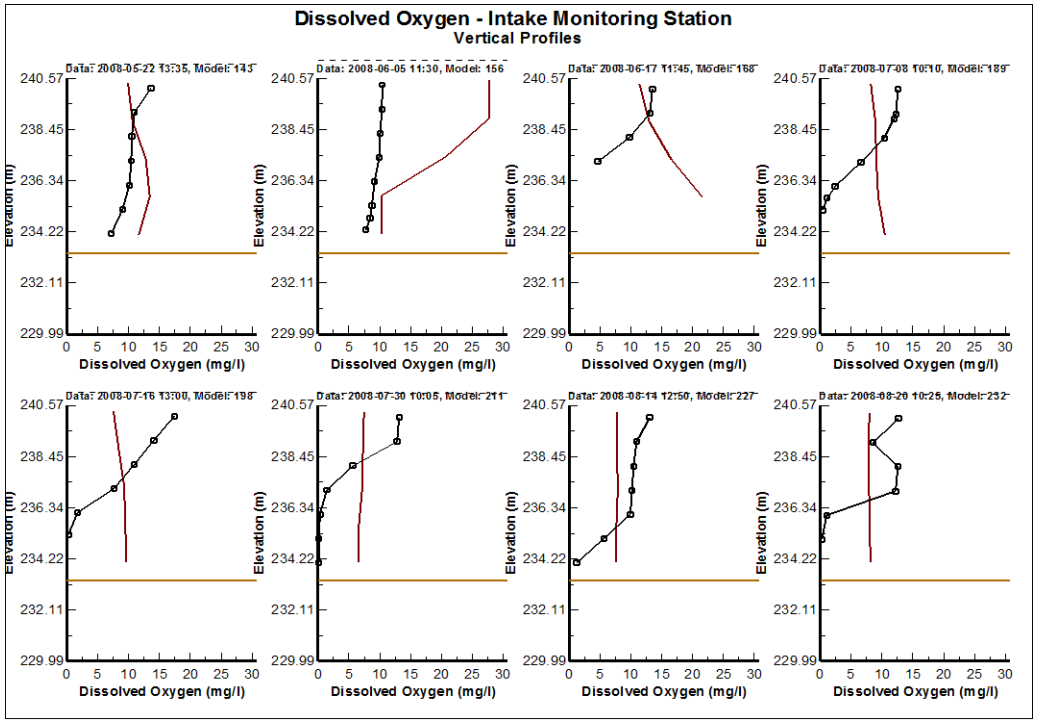


Figure 45: Cyanobacteria - water quality time series calibration results for Dam monitoring station.

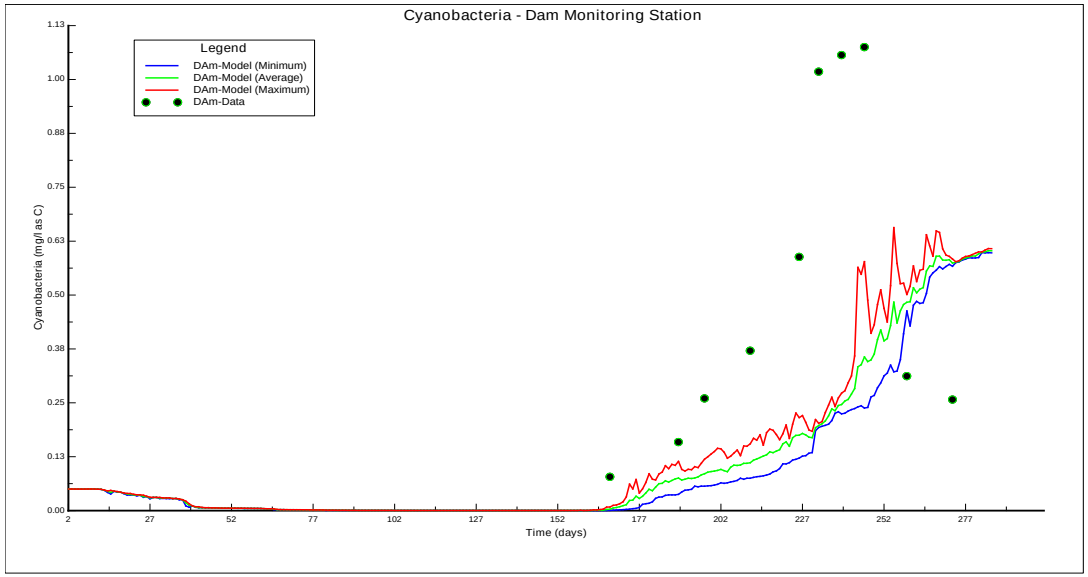


Figure 46: Diatoms – water quality time series calibration results for Dam monitoring station.

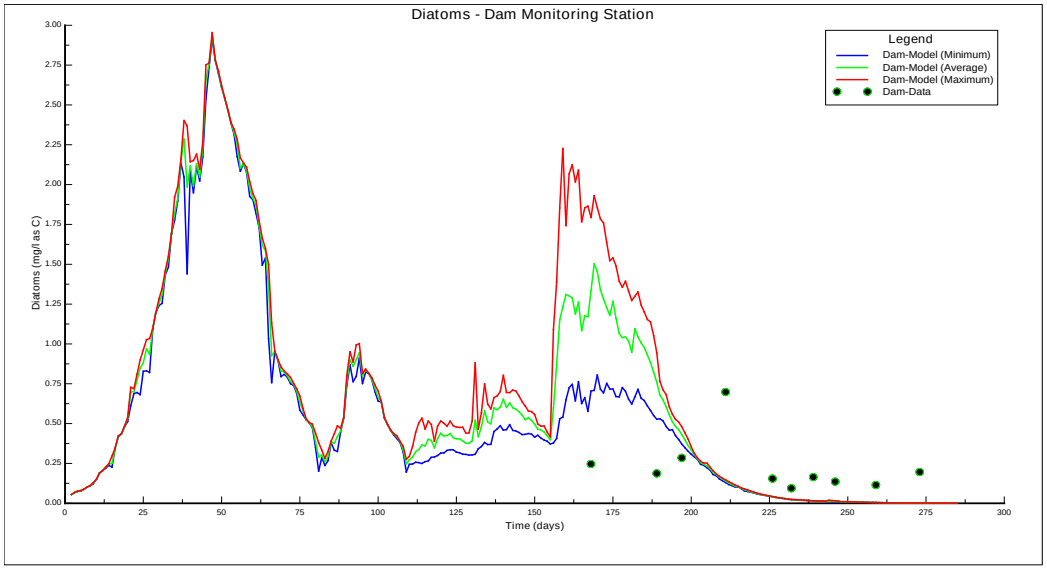


Figure 47: Green Algae– water quality time series calibration results for Dam monitoring station.

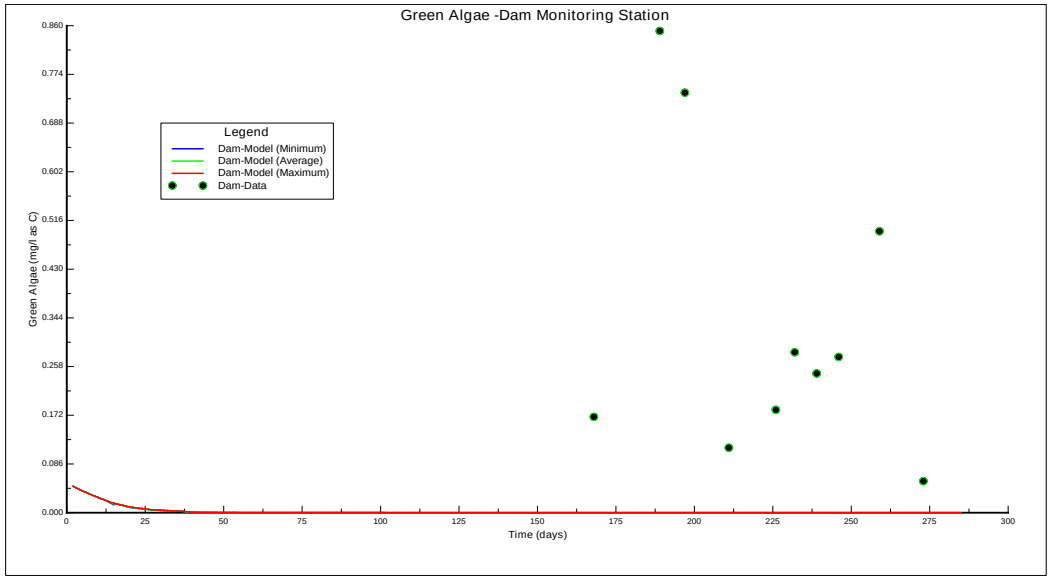


Figure 48: Cyanobacteria– water quality time series calibration results for Marina monitoring station.

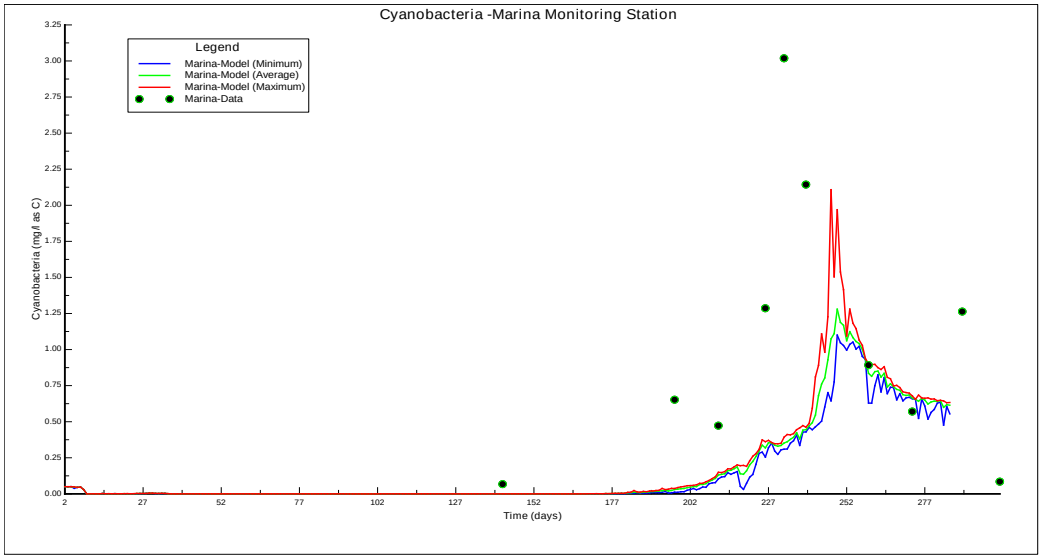


Figure 49: Diatoms –water quality time series calibration results for Marina monitoring station.

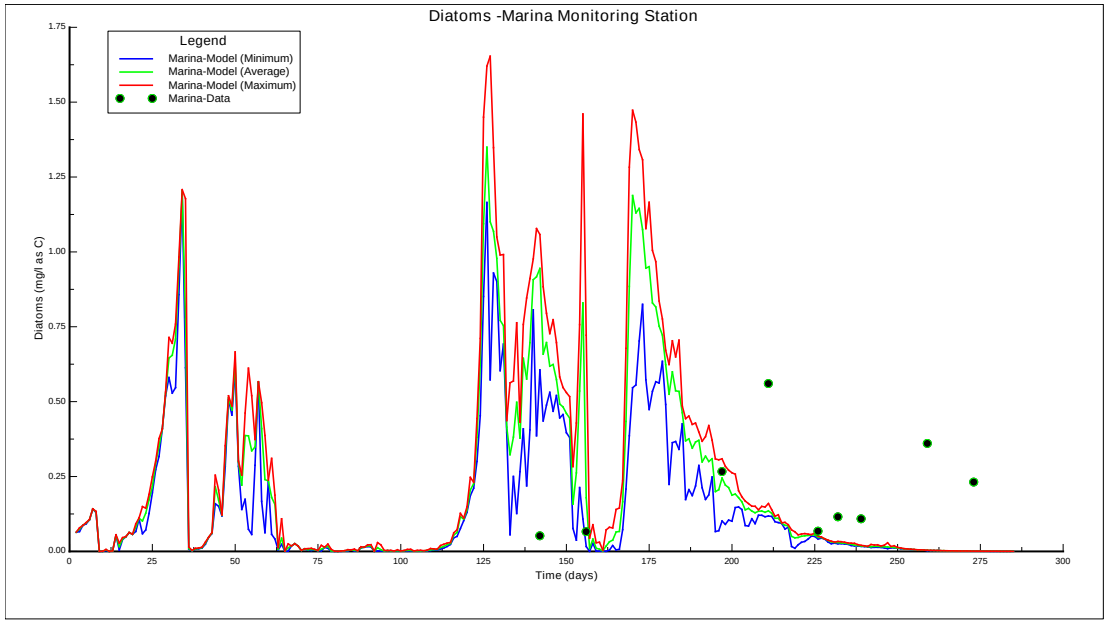
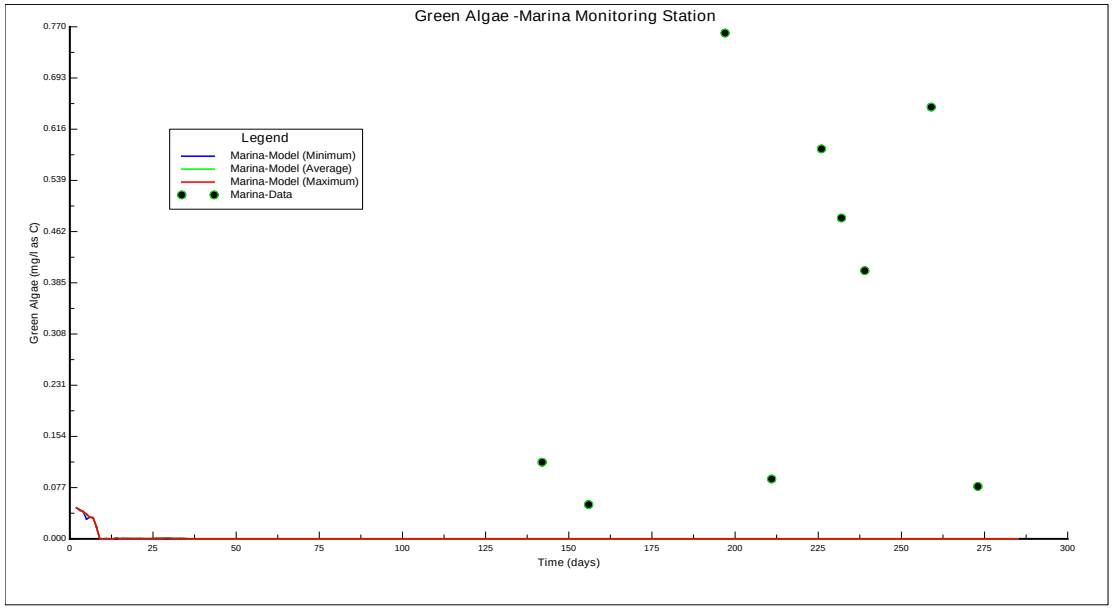


Figure 50: Green Algae - water quality time series calibration results for Marina monitoring station.



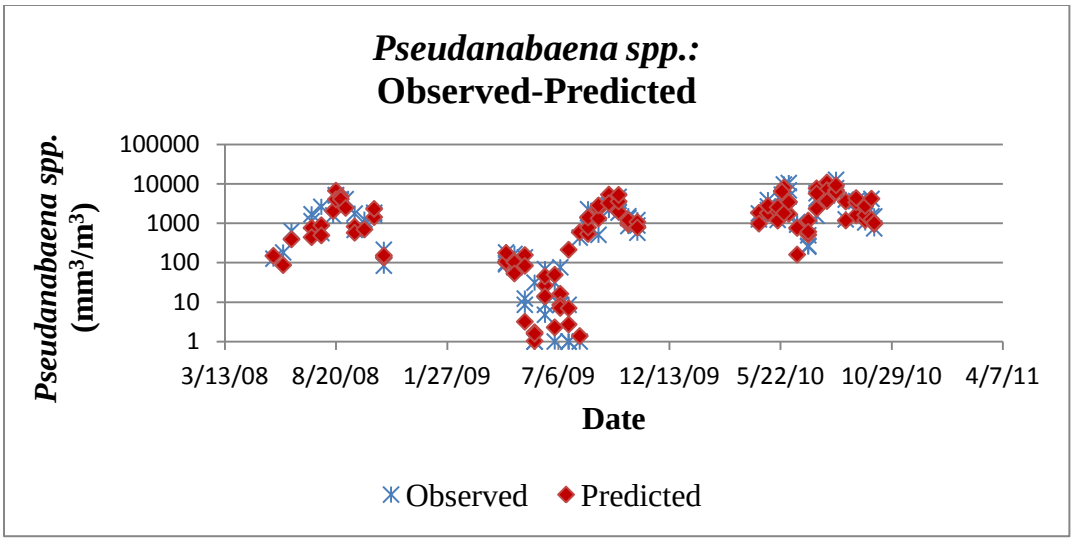


Figure 51: *Pseudanabaena spp*. time series

for training data set: observed versus predicted values.

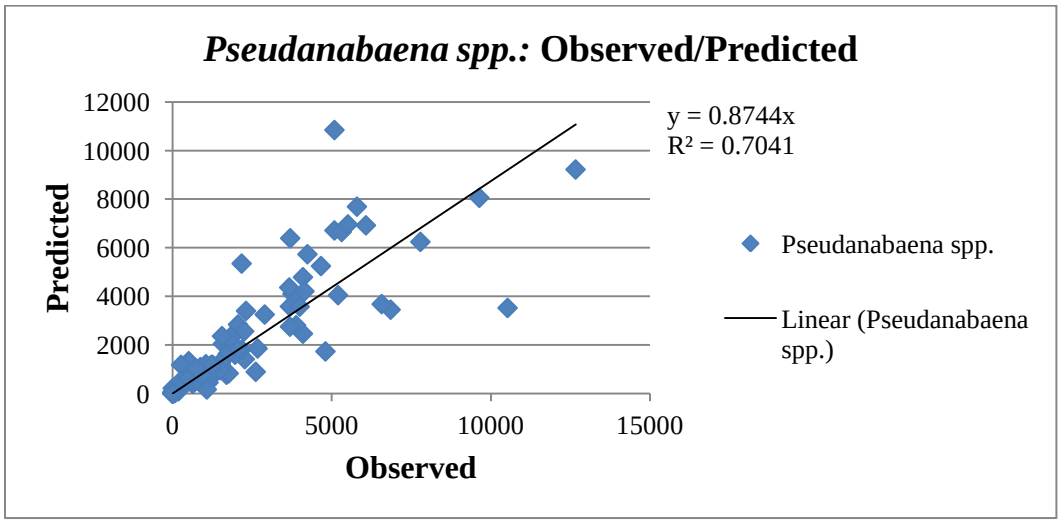


Figure 52: Coefficient of determination

for *Pseudanabaena spp.* training set.

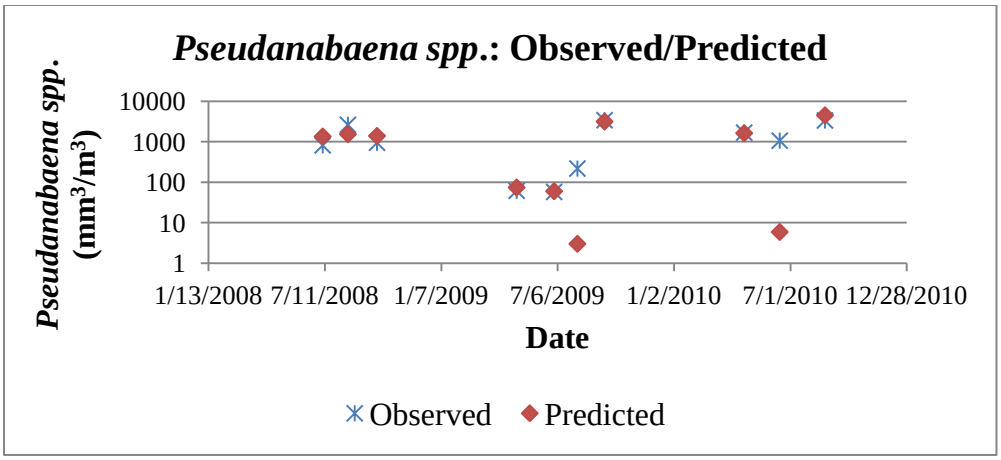


Figure 53: *Pseudanabaena spp*. model validation results for 2008-2010 (ECR)

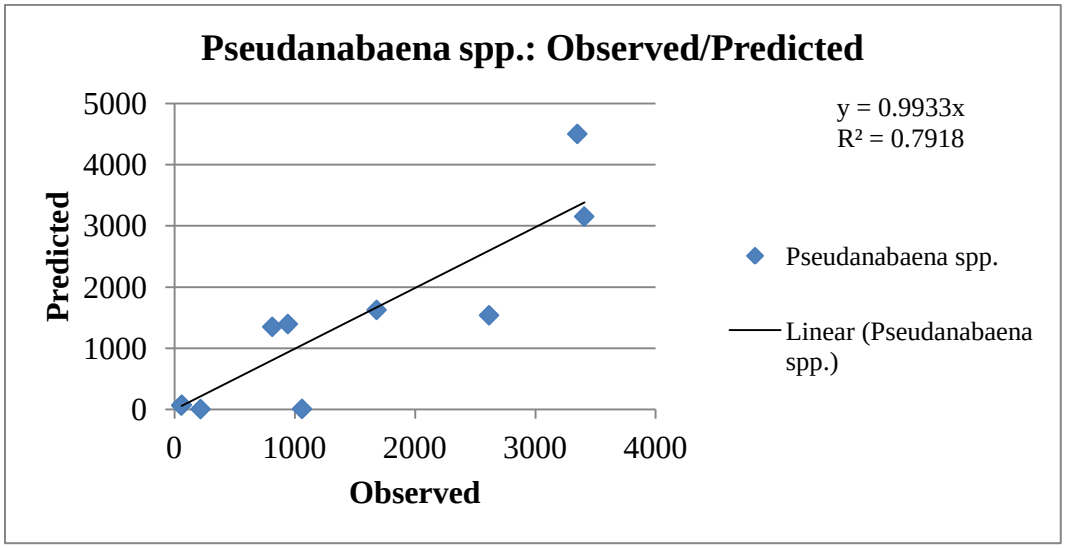


Figure 54: *Pseudanabaena spp.* model coefficient of determination validation results

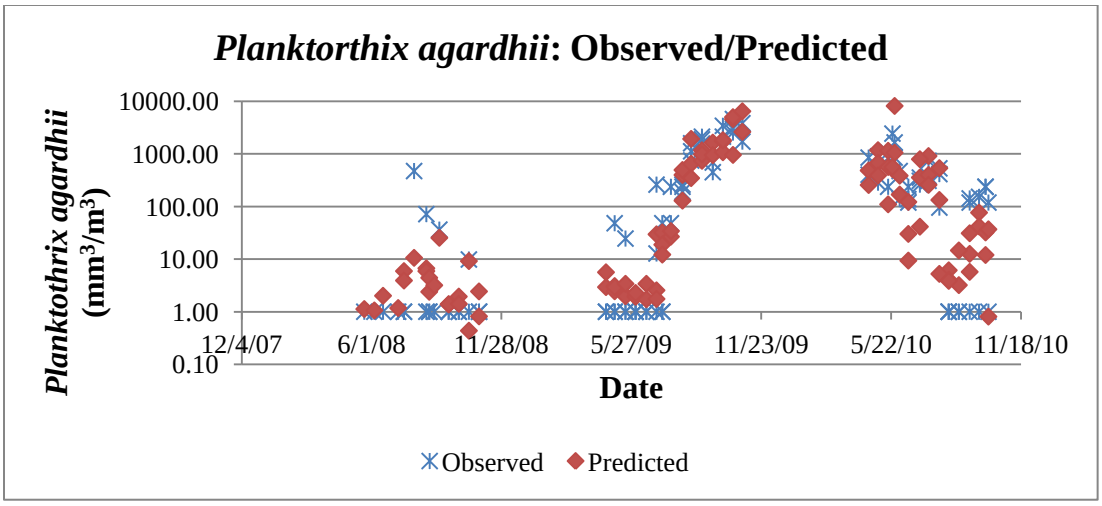


Figure 55: *Planktothrix agardhii* time series

for training data set: observed versus predicted values.

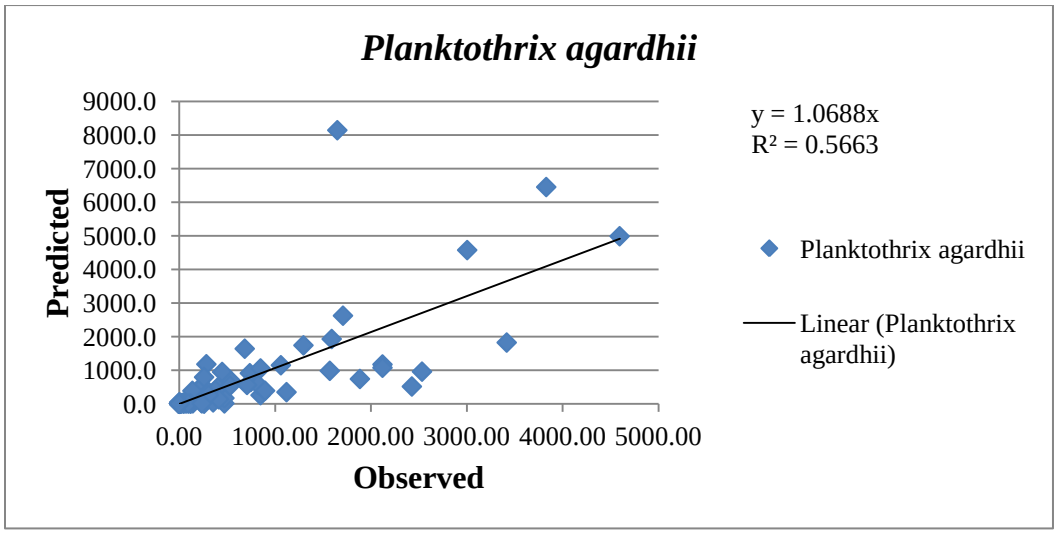


Figure 56: Coefficient of determination for *Planktothrix agardhii*. training set

:

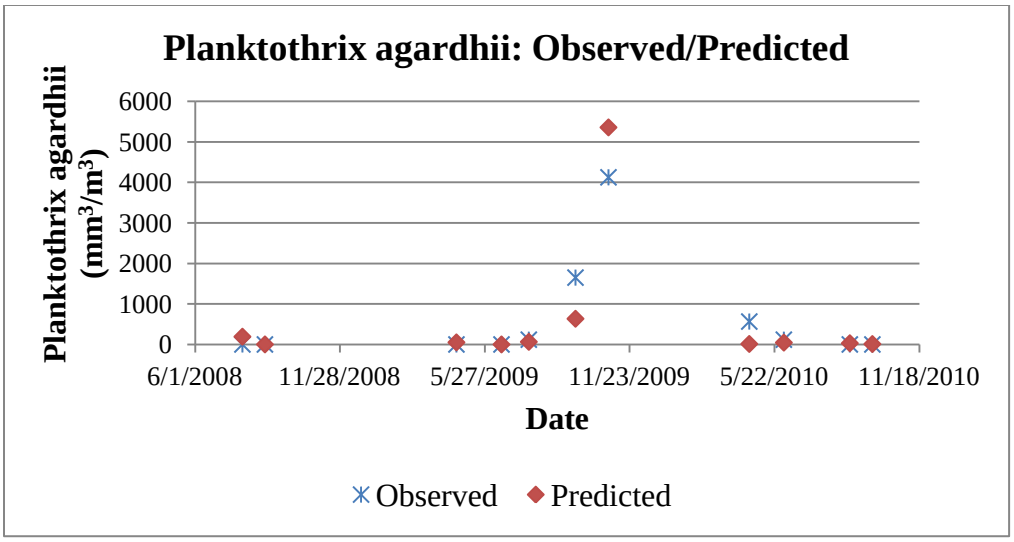


Figure 57: *Planktothrix agardhii* model validation results for 2008-2010 (ECR).

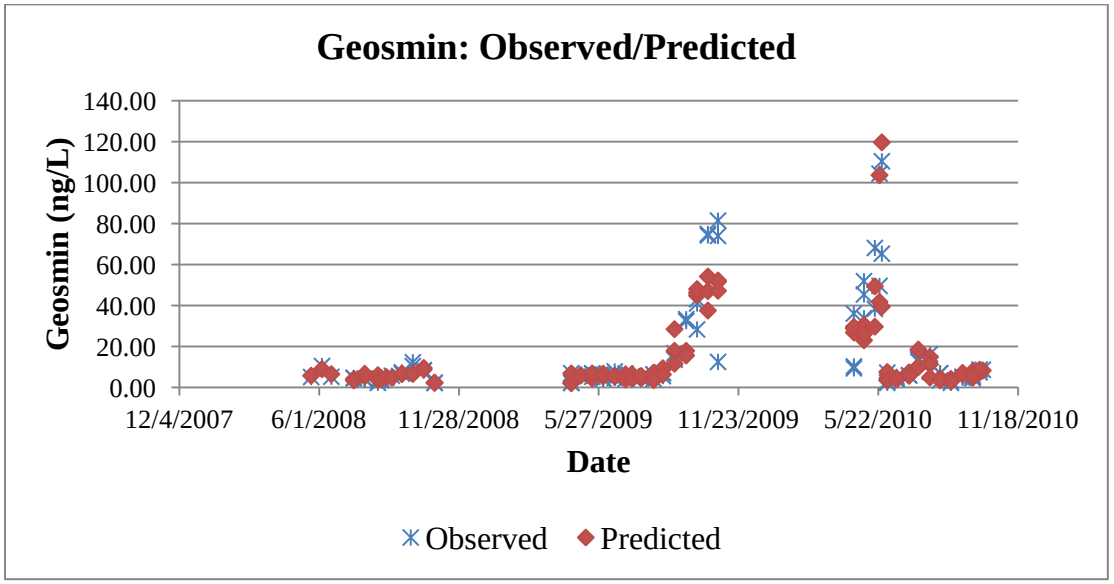


Figure 58: Geosmin time series for training data set: observed versus predicted values

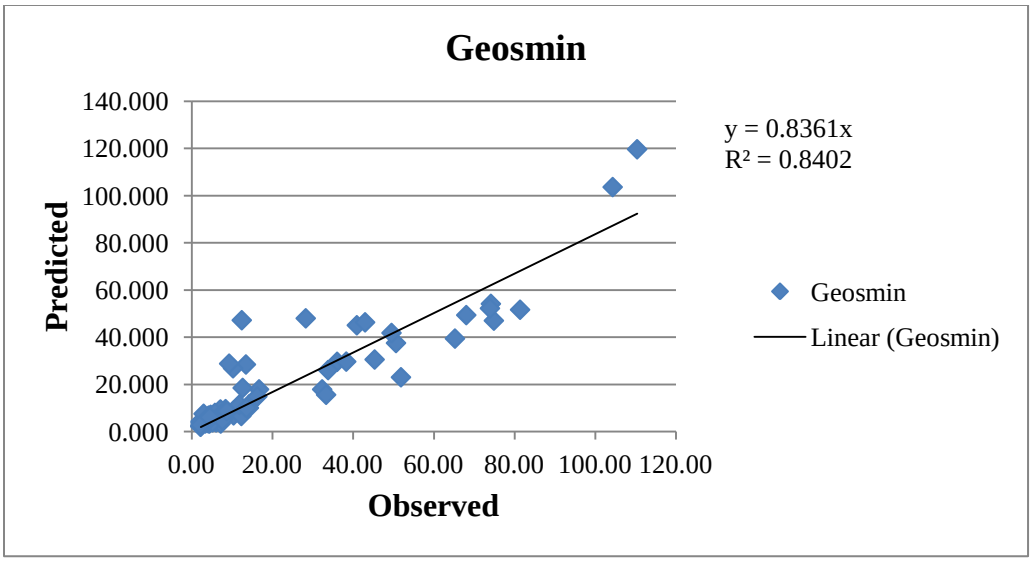


Figure 59: Coefficient of determination for geosmin training set.

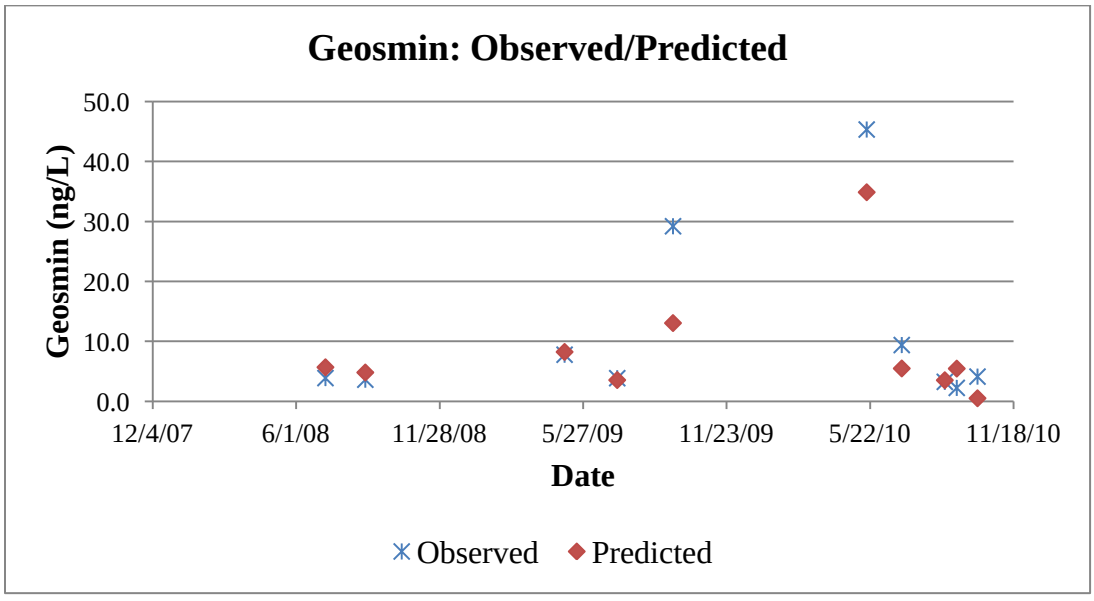


Figure 60: Geosmin model validation results for 2008-2010 (ECR).

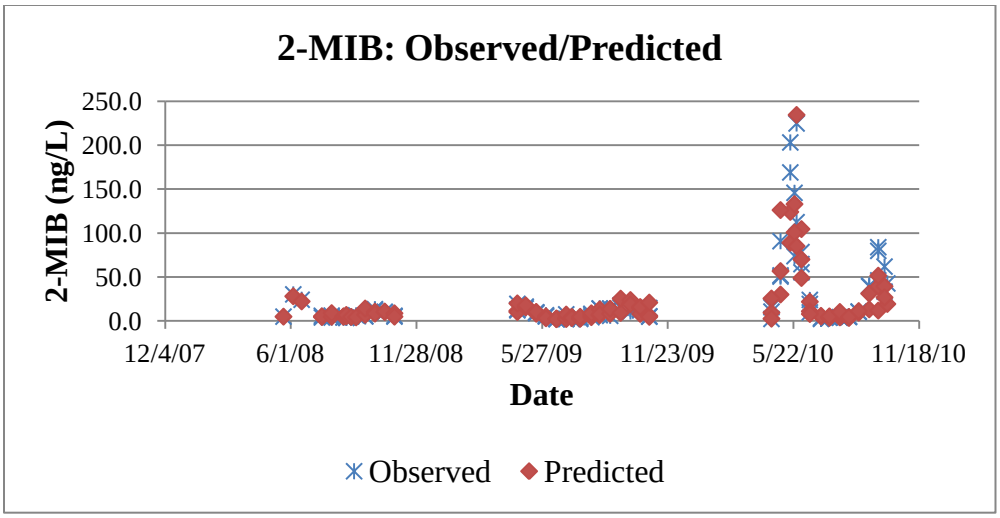


Figure 61: 2-MIB time series for training data set: observed versus predicted values.

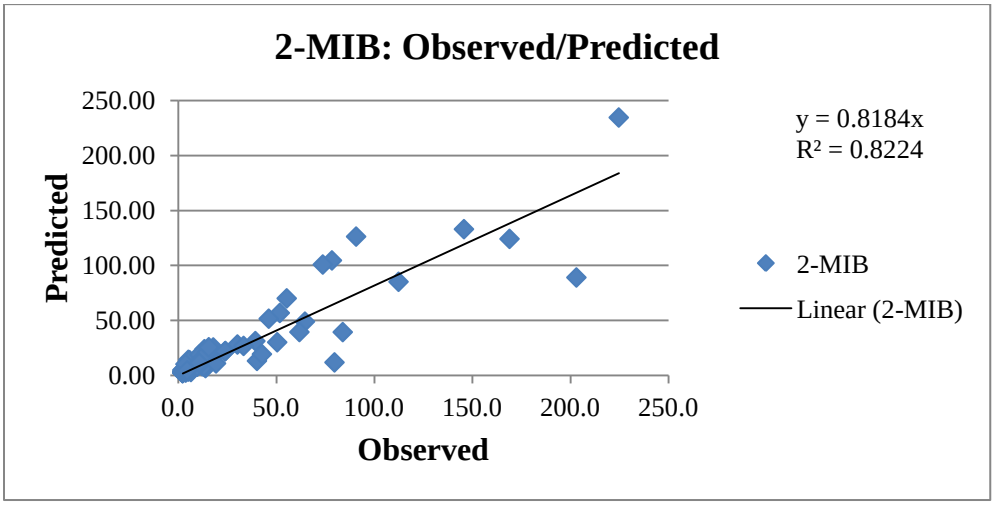


Figure 62: Coefficient of determination for 2-MIB training set.

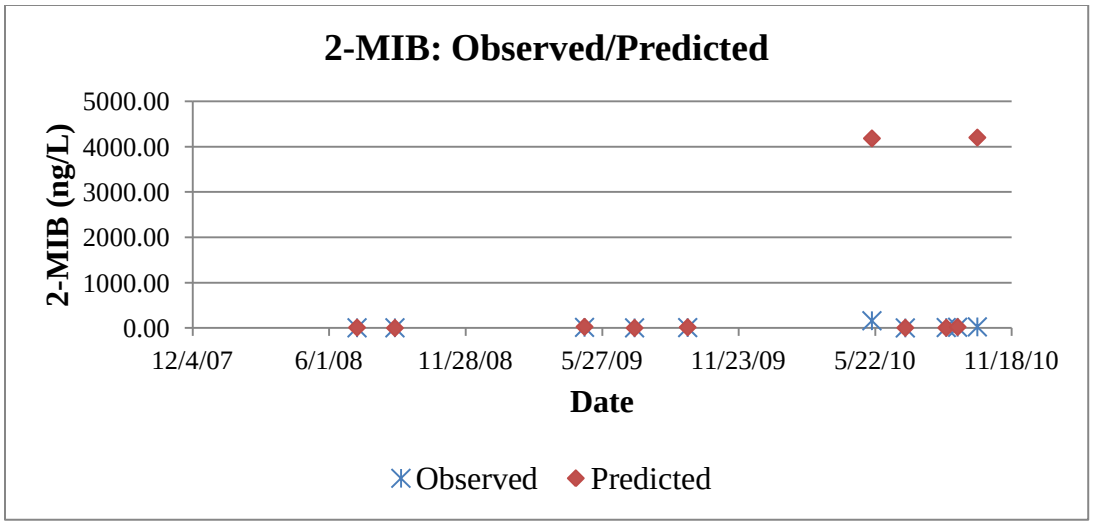


Figure 63: 2-MIB model validation results for 2008-2010 (ECR).

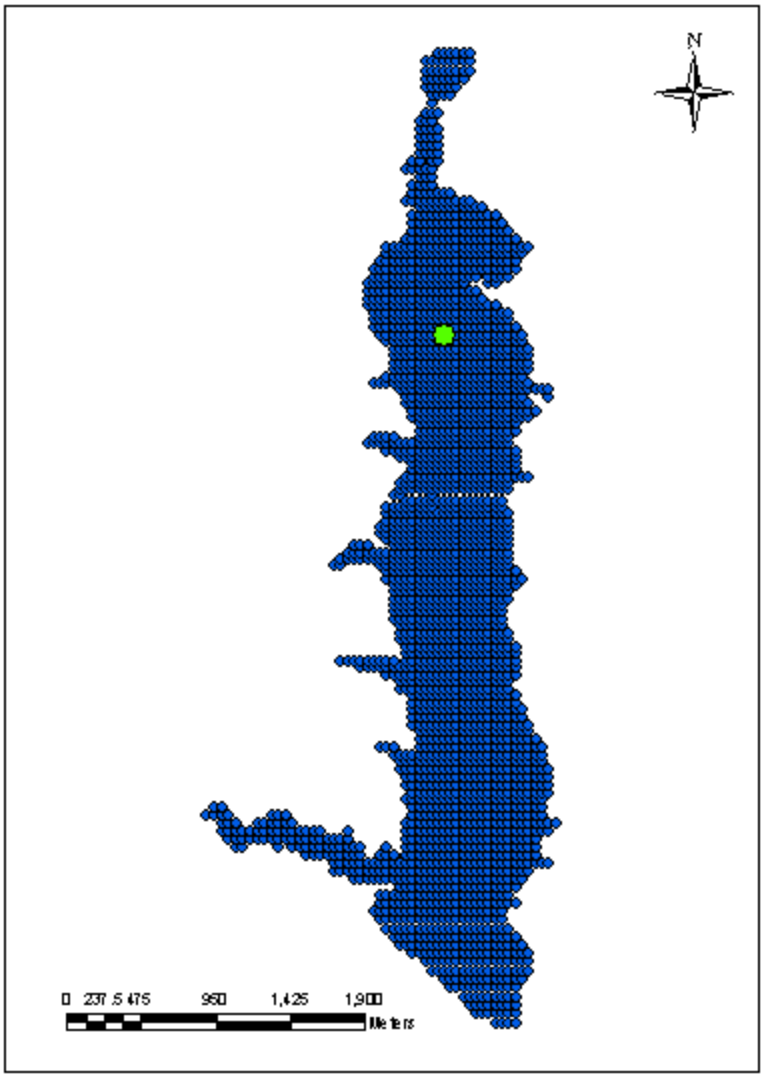


Figure 64: Location of September 30, 2008 water quality measurements (green circular symbol).

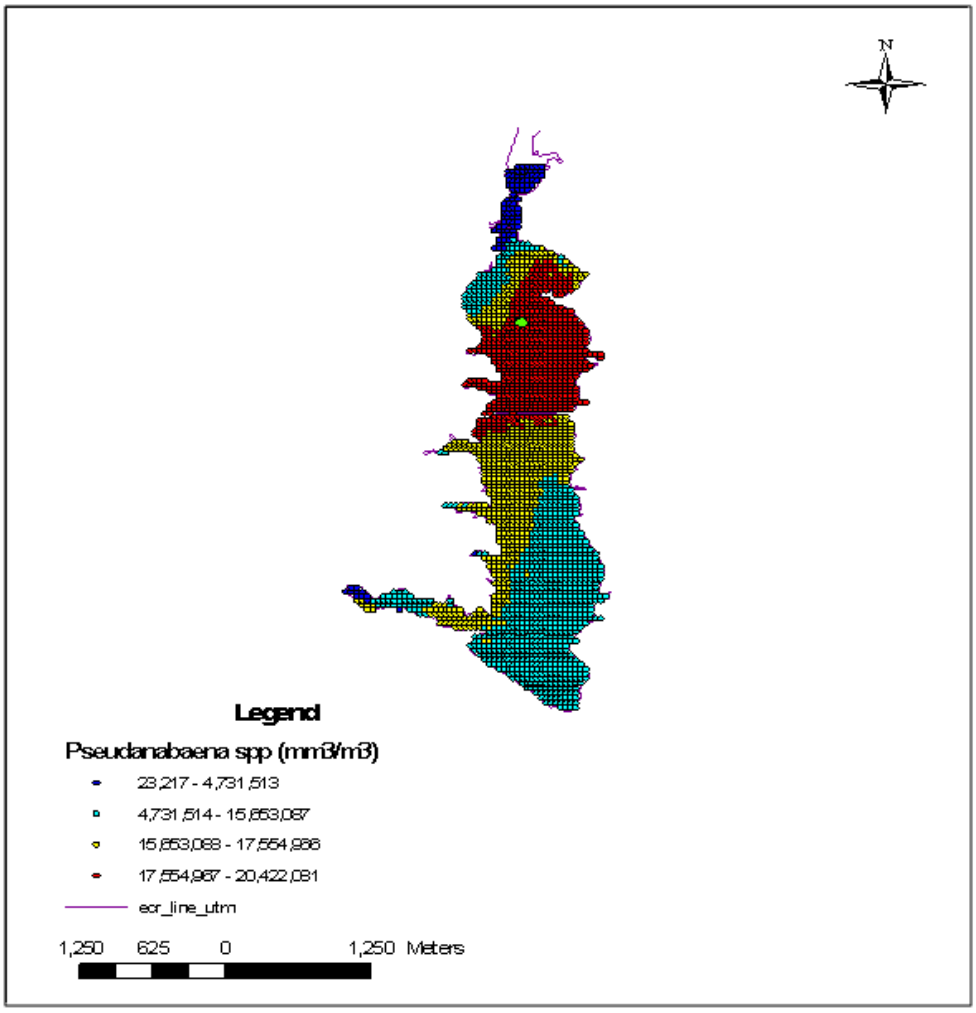


Figure 65: *Pseudanabaena spp*. biovolume prediction for ECR.

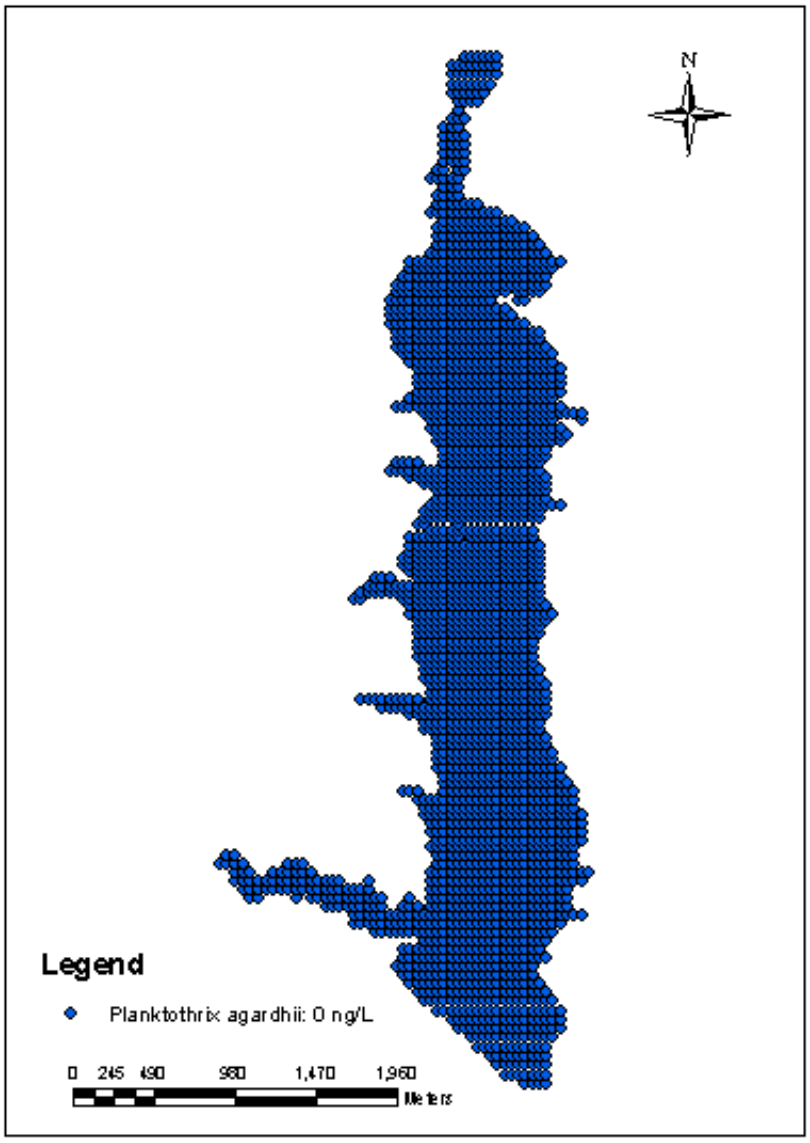


Figure 66: *Planktothrix agardhii* biovolume prediction for ECR.

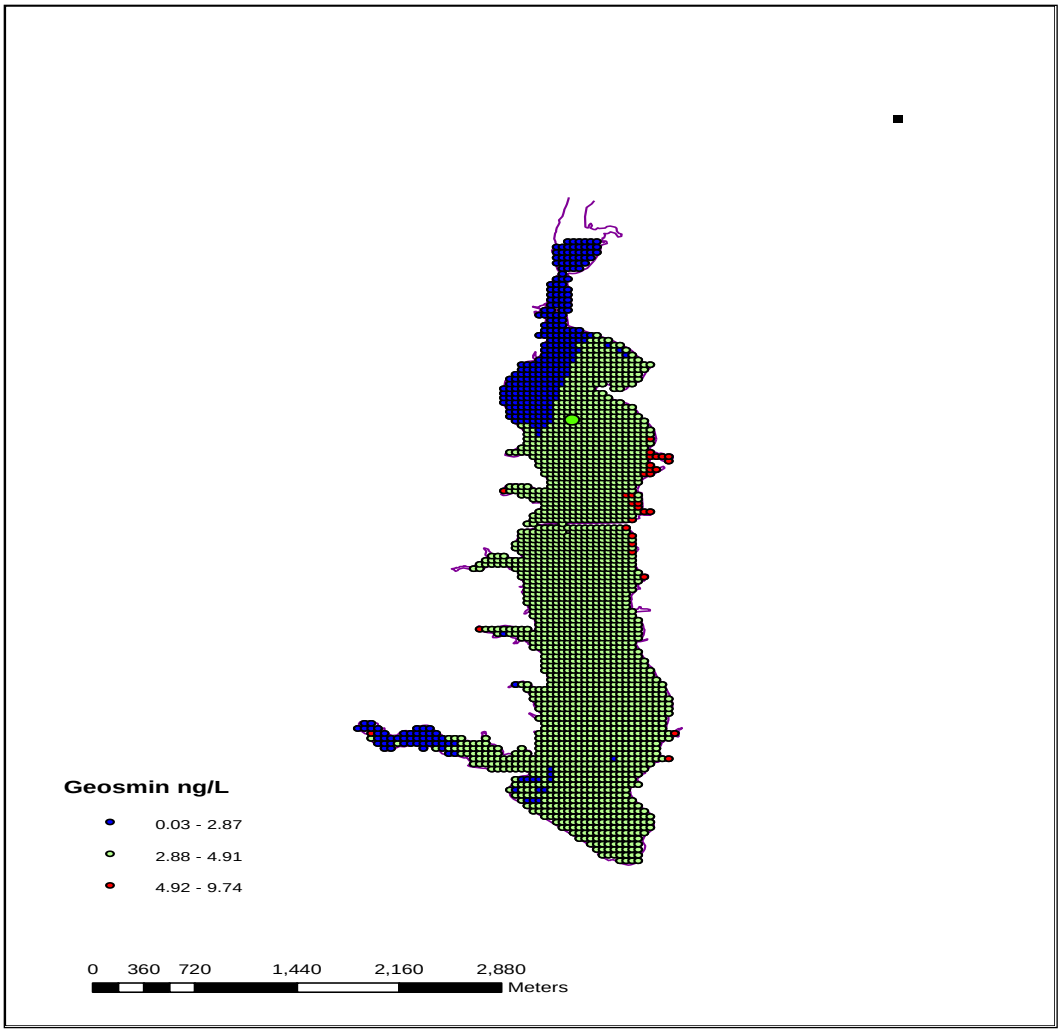


Figure 67: Geosmin concentration prediction for ECR.

Appendix

Taste and Odor (T&O) Compounds

The most often literature described T&O compounds include earthy-musty geosmin, 2-methylisoborneol (2-MIB) and 2, 4, 6 trichloroanisole. Geosmin and 2-MiB are tertiary alcohols, which are present as (+) and (-) enantiomers (Jüttner & Watson, 2007). These alcohols are produced in certain microorganism's cells, and the pathways of their production are constantly under investigation. Geosmin occurs in cyanobacterial cells as two distinct intracellular fractions, i.e., one, dissolved in the aqueous cytosol and two, bound to proteins. Geosmin and 2-MIB are relatively stable to chemical and biological degradation and can exist in the open water in the dissolved form for some time (Jüttner, Watson, 2007).

Presence of these two alcohols in drinking water can have different origins. Significant fraction of T&O compounds are produced by certain groups of pelagic and benthic aquatic microorganisms found in reservoirs, lakes and other surface waters. They are recognized to have strong spatial and seasonal patterns connected with the growth of their producers (Peter, 2008). Besides being formed by aquatic microorganisms, T&O compounds can also originate from terrestrial ecosystems such as drinking water plants and waste treatment facilities (formed during oxidation and disinfection of drinking water or generated in the water treatment distribution system -pipes, bio-films). According to Peter (2008) approximately 200 algal T&O compounds have been identified to date. Their physiological role is still being investigated, but some of them are being described as repellants against pheromones and grazers. Some of these compounds can be biologically active and some are formed within cell of the organisms (algae) and released during cell membrane rapture (death of the organism, grazing by other organisms, cell membrane damage). Another source of T&O compounds are soil aerobic filamentous actinomycete bacteria (*Streptomycetes*), which often are introduced to lakes and reservoirs from snowmelt and terrestrial runoff. Although, actinomycetes production is often considered to play less significant role in 2-MIB and geosmin production, it cannot be ignored as several studies mentioned by Jüttner (2007) i.e. Lake Kasumiguara (Japan),

Lake Schleinsee (Germany) indicated that actinomycetes in some cases could be significant source of odor episodes in aquatic systems.

Existing research has shown that under normal conditions cyanbobacteria (photoautotrophic producer) and actinomycetes (heterotrophic producers) produce very little of their metabolites into the medium and that the excretion of 2-MIB & geosmin is caused by grazing and environmental stressors such as photooxidation, desiccation, water level changes and other (Jüttner & Watson, 2007). Geosmin also can be lysed to water column due to accelerated sedimentation of producing cyanobacterial microorganisms in stagnant waters.

The uncertainty of interactions between species and unknown contribution of actinomycetes are one of the challenges of accurate prediction of T & O events. While contribution of MIB and geosmin from actinomycetes was beyond scope of this research, composition of pelagic algal species and their relation to concentrations of MIB and geosmin were investigated. Although, the selected benthic algal species are within taste and odor metabolite producers, they weren't incorporated in modeling scheme similarly to actinomycetes due to lack of available data.

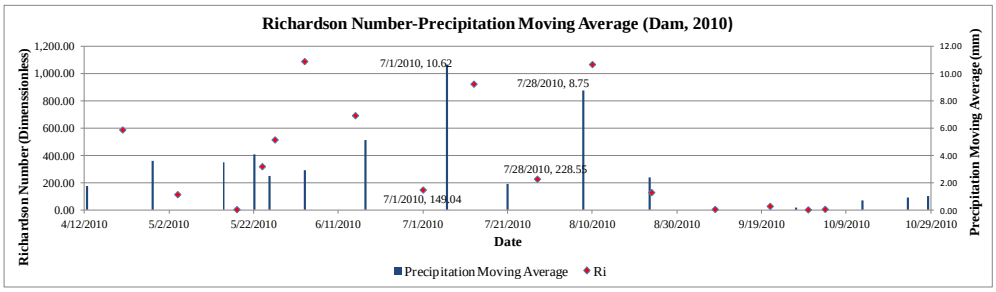


Figure I. Richardson Number in comparison with precipitation moving average for Dam station, 2010.

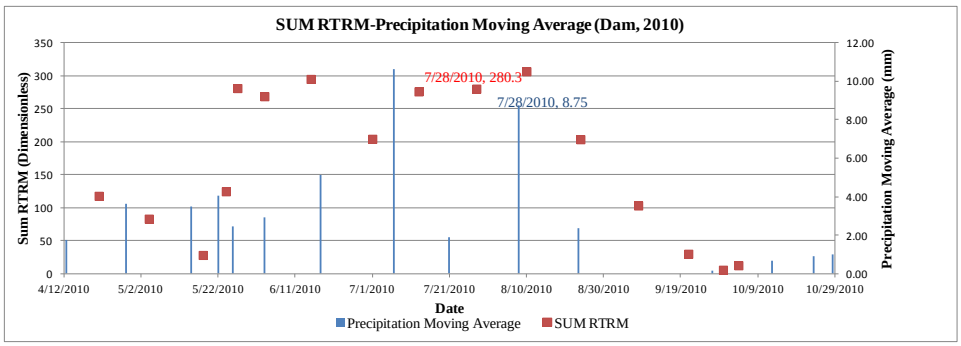


Figure II. Sum RTRM in comparison with precipitation moving average for Dam station, 2010.

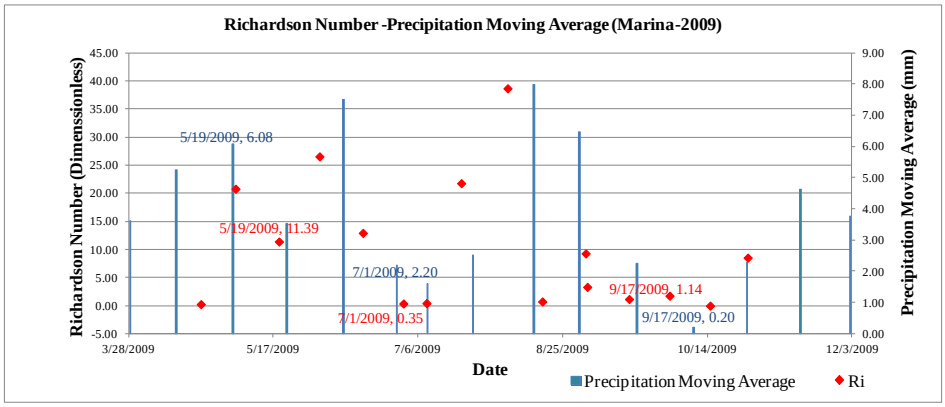


Figure III. Richardson Number in comparison with precipitation moving average for Marina station, 2009.

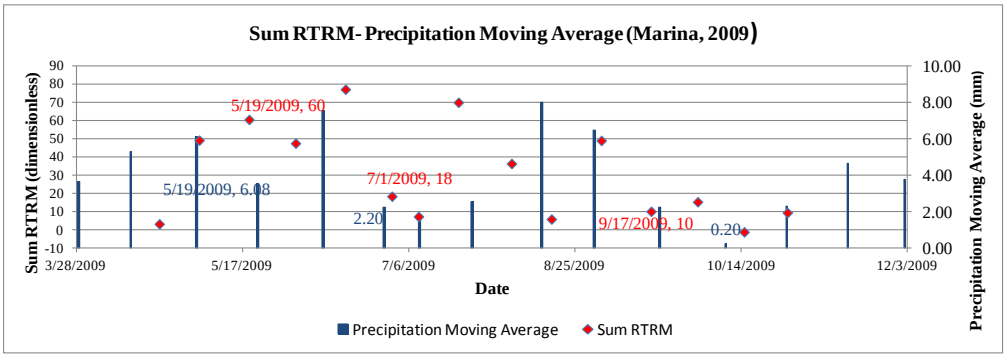


Figure IV. Sum RTRM in comparison with precipitation moving average for Marina station, 2009.

References

Blauw AN, Anderson P., Estrada M., Johansen M., Laanemets J., Peperzak L., Purdie D., Raine R., Vahtera E.(2006). The use of fuzzy logic for data analysis and modeling of European harmful algal blooms:Results of the HABES project. *African Journal of Marine Science*, 28(2): 365-369.

Bright Greg, Cutler Amanda (2002). A biological assessment of the Eagle Creek Reservoir Watershed to diagnose water quality problems. Submitted as an attachment to the October 2002 Draft WMP.

Cerco Carl F. (2000). Phytoplankton Kinetics in the Chesapeake Bay Eutrophication Model. *Water Quality & Ecosystems Modeling*. Volume 1, Number 1-4, 5-49.

Chapra Steven C. Surface water-quality modeling. Long Grove, Illinois: Waveland Press, Inc., 1997. Print

Chen Qiuwen, Mynett Arthur E. (2006). Modeling algal blooms in the Dutch coastal waters by integrated numerical and fuzzy cellular automata approaches. *Ecological Modeling* 199, 73-81.

Cronk Brian C. (2002). How to use SPSS. A step-by-step guide to analysis and interpretation. Second edition. Pyrczak Publishing. ISBN 1-884585-42-6.

Dodds Walter K., Bouska Wes W., Eitzmann Jeffrey L., Pilger Tyler J., Pitts Kristen L., Riley Alyssa J., Schloesser Joshua T., Thornbrugh Darren J. (2009). Eutrophication of U.S. Freshwaters:Analysis of Potential Economic Damages. *Environmental Science & Technology,43(1), 12-19.*

Domingues Rita B., Anselmo Tânia P., Barbosa Ana B., Sommer Ulrich, Galvão Helena M. (2011). Nutrient limitation of phytoplankton growth in the freshwater tidal zone of a turbid, Mediterranean estuary. Estuarine, Coastal and Shelf Science 91 282-297.

Dzialowski Andrew R., Huggins Donald G., deNoyelles Jerry, Jakubauskas Mark, Lim Niang Choo, Beury Jason (2007). Predicting taste and odor events in Kansas Reservoirs – Phase 1. Kansas Biological Survey Report No. 143.

Dzialowski Andrew R., Smith Val H., Huggins Donald G., deNoyelles Frank, Lim Niang-Choo, Debbie S. Baker, Beury Jason H. (2009). Development of predictive models for geosmin-related taste and odor in Kansas, USA, drinking water reservoirs. *Water Research* 43, 2829-2840.

Ernst Mark R., Owens Jennifer (2009). Development and application of a WASP model on a large Texas reservoir to assess eutrophication control. *Lake and Reservoir Management.* Volume 25, Issue 2, p. 136-148.

Galloway Joel M., Green Reed W. (2006). Analysis of Ambient Conditions and simulation of hydrodynamics and water-quality characteristics in Beaver Lake, Arkansas, 2001 through 2003. USGS Scientific Investigations Report 2006-5003.

Ghani Rasli A., Mohamed Azah, Shareef Hussain (2009). ANFIS Aproach for Locating Precise Fault Ponts with Coordinated Geometrics in a Test Distribution System. *European Journal of Scientific Research* ISSN 1450-216X Vol. 35 No.3, pp. 41-473.

Hack Eileen, Tedesco Lenore P., Floress Kristin, Procopy Linda S. (2008). Using social indicator research to enhance watershed education for drinking water resources: Eagle Creek Watershed, IN. *Lakeline* 23.

Hamrick John M. (1992). A three-dimensional environmental fluid dynamics computer code: theoretical and computational aspects.

Hobson Peter, Fazekas Caroline, House Jenny, Daly Rob, Kildea Tim, Giglio Steve, Burch Michael, Lin Tsair-Fuh, Chen Yan-Min (2010). Tastes and Odors in Reservoirs. Water Quality Research Australia (WQRA) Research Report No 73.

Jang Jyh-Shing Roger , Sun Chuen-Tsai (1995). Neuro-Fuzzy Modeling and Control. *Proceedings of the IEEE*, Vol. 83, No. 3.

Journey Celeste A., Arrington Jane M., West Rebecca, Westcott John W., Tuck Ken (2008). Geosmin Occurrence in Lake William C. Bowen and Municipal Reservoir #1, Spartanburg County, South Carolina, 2005 to 2006. Proceedings of the 2008 South Carolina Water Resources Conference, October 14-15, 2008, Charleston.

Jüttner Friedrich., Watson Susan B. (2007). Biochemical and Ecological Control of Geosmin and 2-Methylisoborneol is Source Waters. *Applied and Environmental Microbiology*, p. 4395-4406, Vol.73, No. 14.

Kerr J. P., Thurtell G. W., Tanner C. B. (1967). An integrating pyranometer for climatological observer Stations and mesoscale networks. J. Appl. Meterol. 6:688-694.

Kortmann R. W., Henry D. D., Kuether A., Kaufmann S. (1982). Epilimnetic nutrient loading by metalimnetic erosion and resultant algal responses in Lake Waramaug, Connecticut. *Hydrobiologia* 92, 501-510.

Ligor M., Buszewski B. (2006). An Investigation of the Formation of Taste and Odor Contaminants in Surface Water Using the Headspace SPME-GC/MS Method. *Polish Journal of Environmental Studies*, Vol.15, No.3, 429-435.

Lobugeois Florent (2009). Development of A 3D Hydrodynamic Model For Eagle Creek Reservoir. IUPUI 09/2009.

Malve Olli, Laine Marko, Haario Heikki, Kirkkala Teija, Servala Jouko (2007). Bayesian modeling of algal mass occurrences-using adaptive MCMC methods with a lake water quality model. *Environmental Modeling & Software* 22, 966-977.

Marsili-Libelli S. (2004). Fuzzy prediction of the algal blooms in the Orbetello lagoon. *Environmental Modeling & Software* 799-808.

McGovern Teresa (2006). Lake Water Quality Model with Focus on Cyanobacteria. Water Environment Fundation, WEFTEC 2006.

Munk Walter H., Anderson Ernest R. (1948). Notes on a theory of the thermocline. *Journal of marine research* 1/01, 276-286.

Nielsen Eric J. (2005). "Algal succession and nutrient dynamics in Elephant Butte Reservoir". MS thesis. Department of Civil and Environmental Engineering, Brigham Young University.

Peter Andreas (2008). Taste & Odor in Drinking Water: Sources & Mitigation Strategies. Dissetration. Swiss Federal Institute of Technology Zurich. Print.

Recknagel Friedrich (1997). ANNA-Artificial neural network model for predicting species abundance and succession of blue-green algae. *Hydrobiologia* 349, 47-57.

Takagi Tomohiro, Sugeno Michio (1985). Fuzzy identification of systems and its applications to modeling and control. *IEEE Transactions On Systems, Man, And Cybernetics*, Vol. SMC-15, No.1, January/February.

Taylor William D., Losee Richard F., Torobin Marcia, Izaguirre George, Sass Debra, Khiari Djanette, Atasi Khalil (2006). Early warning and management of surface water taste-and-Odor events. The Awwa Research Fundation.

Tetra Tech, Inc. (2007). The Environmental Fluid Dynamic Code. Theory and Computation. Volume 1: Hydrodynamics and Mass Transport.

Tetra Tech, Inc., (2007). The Environmental Fluid Dynamic Code. Theory and Computation. Volume 3: Water Quality Module.

Tetra Tech, Inc. (2009). Hydrodynamic & water quality modeling report for Lake Lanier, Georgia. Lake Lanier modeling report.

Tillman Dottie H., Cerco Carl F., Noel Mark R.,Martin James L., Hamrick John (2004). Three dimensional eutrophication model of the lower St. Johns River, Florida. U.S. Army Corps of Engineers.Environmental Laboratory ERDC/EL TR-04-13.

Turkmen M., Yildiz C., Guney K., Kaya S. (2009). Comparison o Adaptive – Network-Based Fuzzy Inference System Models Fo Analysis of

Conductor-Backed Asymetric Coplanar Waveguides. *Progress IN Elecromagnetics Research M.*, Vol.8 1-13.

USEPA, New England Region (2002). Office of Water Resources, Rhode Island Department of Environmental Management. Total maximum daily load for dissolved oxygen and nutrients to Mashapang Pond, Rhode Island.

Uwins H. K., Teasdale P., Stratton H. (2007). A case study investigating the occurrence of geosmin and 2-methylisoborneol (MIB) in the surface waters of the Hinze Dam, Gold Coast, Australia. *Water Science & Technology*, Vol. 55 No 5 pp 231-238.

Velo-Suárez L., Gutièrrez-Estrada J. C. (2007). Artificial neural network approaches to one – step weekly prediction of *Dinophysis acuminata* blooms in Huelva (Western Andalucia, Spain. *Harmful Algae 6*, 361-371.

Westerhoff Paul, Rodriguez-Hernandez, Baker Larry, Sommerfeld Milton (2005). Seasonal occurrence and degradation of 2-methylisoborneol in water supply reservoirs. *Water Research* 39, 4899-4912.

Westerhoff Paul, Sommerfeld Milton, Baker Larry (2002). Reducing 2-Methylisoborneol (MIB) and geosmin in the metropolitan-Phoenix area water supply. Guidance Manual.

Winston Bryon (2010). "Reservoir Ageing and associated water quality implications at Beaver Reservoir, NW Arkansas". Dissertation. Department of Geosciences, University of Arkansas. Print

Wong Ken T.M., Lee Joseph H.W., Harrison Paul J. (2009). Forecasting of environmental risk maps of coastal algal blooms. *Harmful Algae 8*, 407-420.

Wyrobek Lee (2010). "The influence of algal toxins and taste–and-odor compounds on water usage in the western basin of Lake Erie". MS thesis. Natural Resources and Environment. University of Michigan.

Zhang Ting, Li Lin, Song Lirong, Chen Wei (2009). Effects of temperature and light on the growth and geosmin production of *Lyngbya kuetzingii* (Cyanophyta). *Journal of Applied Phycology* 21, 279-285.

Curriculum Vitae

Sława Romana Bruder

Education

Master of Science, August 2009 – May 2012

Indiana University, IUPUI, Indianapolis, IN

Major: Earth Sciences with concentration in water quality modeling and hydrology

Thesis: Prediction of spatial-temporal distribution of algal metabolites in Eagle Creek Reservoir, Indianapolis, IN.

Master of Science, October 1992–October 1997

Wrocław University, Wrocław, Poland

Major: Geology with specialization in petrology and mineralogy

Thesis: Conditions for crystallization of Jegłowa quartz deposits with assay of applications in laboratory analysis.

Professional Experience

Graduate Research Assistant, August 2009-August 2011

Department of Earth Sciences, IUPUI, Indianapolis, IN

Highway Engineer Assistant, November 1999-June 2000

Indiana Department of Transportation, Indianapolis, IN

Geologist, May 1999-August 1999

Georgia Oilmen's Services, McDonough, Georgia

Geologist – Internship, June 1997-October 1997

Wrocław Institute of Archeology, Wrocław, Poland

Publications and Presentations

Slawa Bruder, Meghna Babbar-Sebens, Shuangshuang Xie, 2011. Framework for prediction of spatial-temporal distribution of algal metabolites in algal-bloom affected water bodies. Abstarct submitted to the World Environmental & Water Resources Congress, Palm Springs, May 2011.